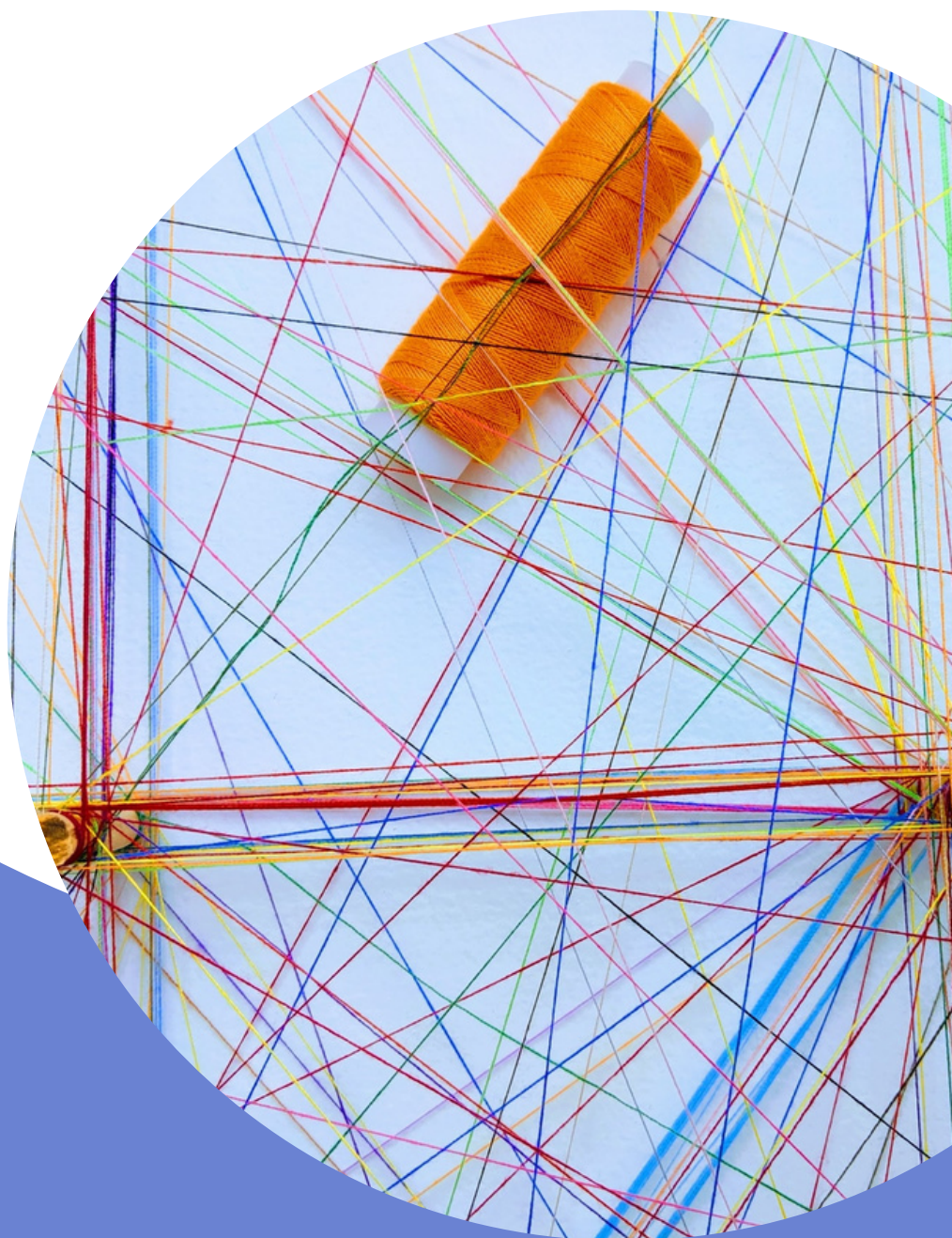


Investigating Research Mobilities: A Methods Resource

Anna Cermakova
Gill Adams
Cathy Burnett
Julia Gillen
Terrie Lynn Thompson
David B. Shannon
Parinita Shetty
Petra Vackova



2024



RESEARCH
MOBILITIES
IN PRIMARY
LITERACY
EDUCATION

**Sheffield
Hallam
University**

Sheffield
Institute
of Education

Lancaster
University



UNIVERSITY of
STIRLING



CONTENTS

01

INTRODUCTION

03

QUALITATIVE APPROACHES TO INVESTIGATE TEACHERS' ENCOUNTERS

Introduction	07
Data generation procedures	08
Analysis	13
Methods combined	31

02

CORPUS LINGUISTICS

03

Introduction	34
Data collection	36
Analytical approaches	45
Analysis of Research Mobilities	55

SOCIOMATERIAL ETHNOGRAPHY

Introduction	67
Selection of cases	75
Data collection	81
Analytic framework	83
Analysis	86

04

CONCLUSIONS

05

Ethics	98
Reflections	99
References	100

Introduction

This methodological resource outlines a suite of methods used as part of the *Research Mobilities In Primary Literacy Education* project (2022-2024, ESRC ES/W000571/1) conducted by Prof. Cathy Burnett (Sheffield Hallam University), Dr Gill Adams (Sheffield Hallam University), Prof. Julia Gillen (Lancaster University), Dr Terrie-Lynn Thompson (Stirling University), Dr Anna Lindroos Cermakova (Lancaster University), Dr David Shannon, Dr Parinitha Shetty and Dr Petra Vackova (all at Sheffield Hallam University) .

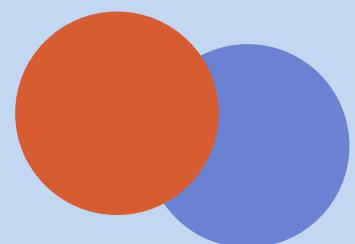
The project aimed to explore how research linked to literacy education moves to, among and around primary teachers. We were also interested more generally in the mobilisation of research - in the **complex networks of individuals, organisations, texts, technologies, media and algorithms and their role in research mobilities**. Another aspect we were interested in is how the “message” of a particular piece of research changes as it moves. For further information on the project and its findings see our book ***Research Mobilities in Primary Literacy Education: How Teachers Encounter Research in An Age of Evidence-Based Teaching*** (published by Routledge and freely available online in 2024). We hope our findings will be of value and interest to researchers, educators, school leaders, policy-makers, literacy charities, teacher educators, and educational consultants.

This resource offers our thinking about the methods we used and is aimed at researchers of research mobilities and beyond. We also hope that, more broadly, our methodologies will be of interest to researchers in the field of knowledge mobilisation but also social sciences researchers in the broader sense, whether they are interested in the methodological assemblage or any of the methods on their own. Most of the examples we use here to illustrate how the methods work and/or how we implemented them are from the *Research Mobilities* project; however, they are chosen so that they are of a more generalisable nature rather than too project specific.

The task of researching research mobilities is impossible to clearly delineate. The numerous actors that participate in the mobility network combine in both expected and unexpected ways. The ideas of '**research mobility**' and '**research**' themselves are extremely slippery notions and thus the borders we placed around our area of enquiry were constantly challenged and adjusted. When we asked teachers to talk about 'research', they would often talk about things that were not research as we understood it, even if their origins might have been.

From the outset, we want to make the point that, in presenting our methodology, we are aiming to show the variety of paths we have tested and taken. The paths were not straightforward, not least due to time constraints of the project. In developing our methodology, we wanted to find ways of engaging with this slipperiness. John Law and Annemarie Mol's ideas about mess and multiplicities were key (Mol 2003, Law & Mol 2002, Law 2004). **Law** (2004, p. 2) sums up the disconnect between the messiness of life and the ordering and organising functions of social science in the opening to his seminal book, ***After Method: Mess in Social Science Research***:

Parts of the world are caught in our ethnographies, our histories and our statistics. But other parts are not or if they are then this is because they have been distorted into clarity. ... If much of the world is vague, diffuse or unspecific, slippery, emotional, ephemeral, elusive or indistinct, changes like a kaleidoscope, or doesn't really have much of a pattern at all, then where does this leave social science? How might we catch some of the realities we are currently missing.



One of the central notions in our thinking is that methods do not just capture realities but help to shape them. Methods (and tools) draw boundaries around what is possible or plausible and what is considered relevant. They **create presences and absences**.

Our approach was to use **three sets of methods** which could be seen as offering **different vantage points** but which all, in effect, bounded research mobilities in different ways. We expected these to enact different kinds of realities that we could put in dialogue with one another to enrich our understanding of research mobilities. The methods allowed us to see the complexity, divergence and many times they were disrupting each other. Our methods included:

01 Qualitative approaches

to investigate teachers' encounters with literacy research with the aim to capture what they encountered as well as insights into their experiences and perspectives.

02 Corpus linguistics

to explore public discourses about primary literacy education research in mainstream media, specifically newspapers; and social media, specifically Twitter (now X).

03 Sociomaterial ethnography

to trace the movements and translations of 'pieces of research', such as research articles or projects, drawing on a combination of network ethnography and controversy mapping.



With these methods we aimed to answer four **research questions**:

- **How does research move to and among teachers?**
- **Which literacy research is circulating?**
- **How does research move? Do meanings shift as research moves?**
- **How can we promote critical engagement with breadth of research for primary literacy education?**

We would like to stress we did not attempt a methodological triangulation and/or single comprehensive overview of research mobilities in primary literacy education, or indeed to sketch a range of different ways in which research, literacy and teaching manifest. In fact, we are not convinced that such an account is possible. Instead, the assemblage of diverse set of methods reflected - and enacted - different perspectives to trace connections, intersections and interruptions between these.

This resource is organized as follows: Sections 2 to 4 explain the three sets of methods we used in more detail and Section 5 briefly outlines some of the ethical considerations and offers reflections.



Qualitative approaches to investigate teachers' encounters

Gill, Parinita, Petra and Cathy

Introduction

We aimed to work alongside teachers as far as possible to understand how they encountered research in their everyday lives. Being mindful of their workloads, we wanted to find ways for them to **record their everyday encounters with literacy research** that were least intrusive but would, at the same time, provide opportunities for them to discuss these records in small groups with the research team.

Teachers were recruited to the project during the academic year 2021/2022. Invitations were circulated to a range of national and local literacy and teacher research networks and groups via email lists but also through social media and newsletters. Our recruitment materials called for teachers with a particular interest in the teaching of literacy/English from different types of primary schools and areas of England.

Our initial approach invited teachers to work longitudinally through a series of activities beginning with a workshop that introduced the project and elicited teachers' stories of encounters with literacy research in their everyday lives. Between March and November 2022, twenty-one teachers participated in several workshops, focus groups, and interviews. All these were held online; they were all recorded and transcribed.

Data generation procedures

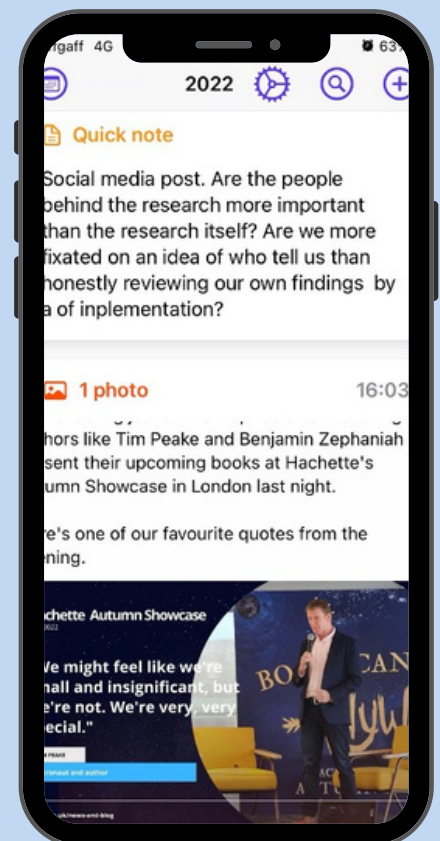
We generated two main types of data:

- **teachers' lifelogging;**
- **transcriptions of the online meetings.**

These datasets are multimodal, heterogenous, and unstructured, which posed challenges for the subsequent analysis. Lifelogging enabled teachers to capture their encounters with primary literacy education research. This provided a record of the research that they noticed, remembered and were able to document. In the online meetings, we discussed these logs, providing teachers with an opportunity to reflect on their encounters and allowing us to gain rich insights into how research reached them, where and when they accessed it, and how these experiences came into relationship with their everyday lives.

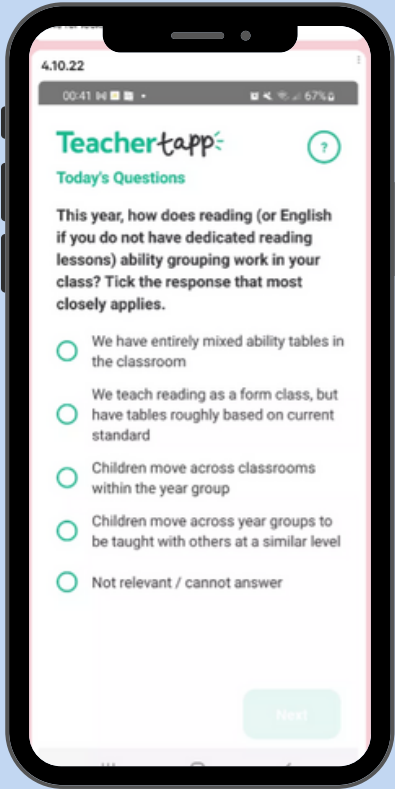
Lifelogging

Our approach to lifelogging builds on Frigo's (2016) work, where **"lifelogging as ethnography attempts to accommodate the elements of mess and chaos, the overflow and complexity of our contemporary reality"** (p. 144), offering an approach that enables individuals to create their own framework "to gather, process, and retrieve information" (p. 143). This enabled teachers to creatively direct the development of this approach and adapt it to fit unobtrusively into their busy lives.



In addition to providing a means to hold on to research encounters that might otherwise be taken for granted and share them with us, it facilitated teachers' **'noticing'** (Mason 2002), offering opportunities for teachers to "better sense and understand the complex media-generated landscape" (Frigo 2016, p. 139) which, together with other material encounters, mobilise research. In order to practically support the teachers' focus on the lifelogging activity, we sent weekly reminders over several weeks.

We encouraged teachers to explore digital media that were readily available; they were creative and adopted a range of tools and strategies in their approaches to lifelogging - they **took photos, made notes, saved screenshots on their phones and laptops, recorded emails and social media trails, used notetaking, both digital and "pen and paper", discussion board apps,** recording at odd moments during the day or at regular intervals during the week according to what worked best for them.



Some of these logs were simple screenshots including elements of multimodality, see the examples.



Some were pictures, some were combination of screenshots and notes, and some were notes, for example in a diary format as the extract below. Lifeloggings were then discussed in focus groups and interviews.

27/5/22 - joined Reddit and some educational 'blogs' from there. 'Educational Research' gives no results.

30/5/22 - on Reddit, it's all anecdotal. Nothing pertaining to evidence or research.

8/6/22 - at work. Wondering if any researchers combine with providers (eg Twinkl) to produce resources, lesson plans or unit with embedded research findings throughout - and how valuable it would be.

The reading fluency article link (in a weekly email from Mary Myatt) took me to



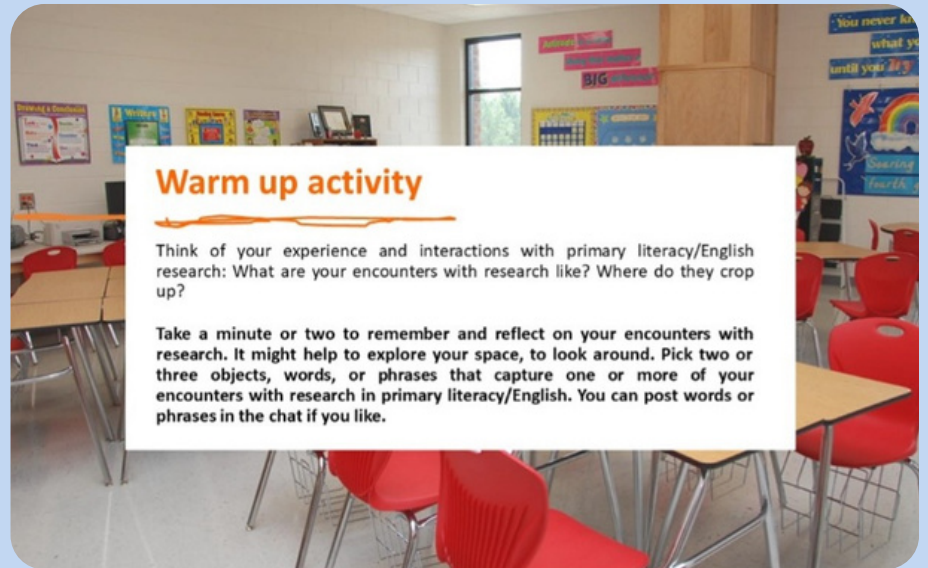
Education
Endowment
Foundation



EEF blog: Shining a spotlight on reading fluency

Workshops

In the workshops, we discussed what “encounters with literacy research” might actually be, for example, how research was being used, what kind of school research-informed initiatives the teachers came across, what were their memories of research previously accessed, who/what was the source of their research.



We encouraged teachers to explore the spaces and objects around them as they reflected on their encounters. The discussions focused on linking ‘research’ to teachers’ own classroom enquiry and practice. The teachers were invited to reflect on **moments when research was notably present or absent**, how research linked to resources or schemes and their use of social media. The workshops followed a similar structure. At the beginning we revisited ethical considerations and consent forms.

Discussion

- Thinking with the 3 objects, words, or phrases you shared with us earlier what is research in primary literacy/English to you?
- What is your experience with research?
- How do you engage with or use research in your work? Or do you not engage with research and why?

After brief introductions, a warm up activity followed, see example above. We then briefly introduced the **Research Mobilities project**; this was then followed by discussion, in which teachers shared experiences of the lifelogging of research encounters (see left the discussion prompt).

Focus groups & interviews

After several weeks of logging, we organised dedicated focus groups and interviews for teachers to share insights from their logs. Logs were used as prompts to explore connections, what interested them and their experience of accessing, logging, sharing and conducting research. Further prompts to facilitate the discussion were used, see below.

What we'd like you to do next is to create a visual representation of how you see research moving to and from you, that thinks across your encounters. Again, you can use any method you choose to record this – you might want to use objects, a big piece of paper, any materials you have nearby, a web tool (mural, padlet, jamboard...). (10 minutes)

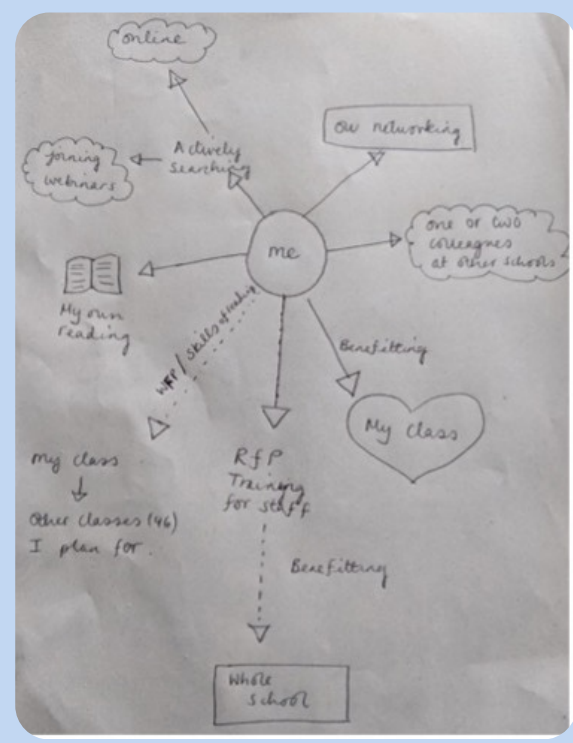
Sharing research encounters – open discussion, asking participants to share one of their records of a literacy research encounter or their visual representations (screen share or hold up to camera). Can you describe the encounter? What is it about this particular encounter that strikes you?

This stage provided an opportunity for teachers to reflect on what they had gathered. As Frigo notes, teachers found “the very experience of sampling and stowing” enriching, which easily prompted reflection (Frigo 2016, pp. 152–153). Finally, after several more weeks of lifelogging, we invited teachers to participate in an interview or focus group where we discussed their encounters, asked them to select and reflect on two or three logs, on their experiences of accessing, engaging with or conducting research, and on the English/literacy topics that cropped up in their encounters.

Adaptations to the initial project plan

Multiple pressures on teachers meant that we did not recruit the planned numbers of teachers to the project in the first round. Building on what we had learned during these early interactions with teachers, and in response to these circumstances, we adapted our initial methodology, planning participation via a single focus group structure. The multiple demands on teachers' time, particularly given the intensive efforts at 'catch-up' in this period post-COVID lockdowns made scheduling a series of meetings difficult. This alternative method aimed to **decrease barriers to participation** and provide opportunities to teachers who may have been unable to participate over a prolonged period.

We were able to reach new participants who offered us fresh insights, using the space to reflect on their experiences, priorities and desires for teachers' engagement with research. These focus groups introduced the project briefly, discussed ethics and the implications of participation, then included a mix of questions and activities that encouraged participants to reflect on their encounters with research. We, for example, asked teachers to create a visual representation (see the paper sketch on the right) of how they saw research moving to and from them and share these with us. Some of these single focus groups effectively became interviews when one or more teachers who had booked spaces were unable to participate.



Data generation: brief reflections

The project itself became a mediator in the movements of research for teacher participants. For example, some teachers talked about the effect that lifelogging had on them, saying that the act of noting an encounter focussed their attention. One had put a literacy podcast on her to-listen list previously and the lifelogging encouraged her to listen to the episode, discover she enjoyed it and list some more. Other participants mentioned research sources or mediators that others subsequently followed up. The methods we used brought together participants who may otherwise never have met, opening up different possibilities and connections. Others were prompted to make time to actively reflect on their research encounters. In these and other ways, participation in the project and the particular methods that were used helped to shape teachers' encounters with research. Part of the project plan was for some participants to become involved in a later stage of the project, developing resources to **support teachers to incorporate research into teaching**; this was another way that the project influenced teachers' engagement with research.

Difficulties in recruitment and participation, likely in part due to the particular social and political circumstances of teachers' working lives, affected the project. Some data generation activities, intended for groups of teachers, effectively became one-to-one interviews, limiting opportunities for sharing perspectives but enabling more in-depth focus on individual's experiences.

Analysis

Tools

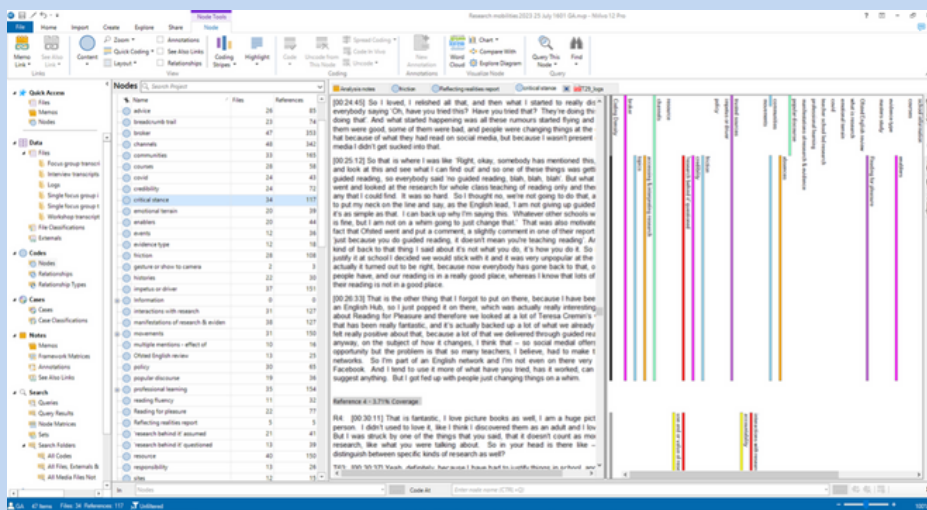
Excel and NVivo

In order to attempt to organise the vast amount of detailed unstructured information we used two tools.

	specific topic	source item name	source type	Source of research, etc. if given, identifiable	Origin of source type	Broker name	Broker type	Channel type	Any other intermediaries
1	reading								
2	interventions	N/A	research paper	N/A	N/A	N/A	university	course	
3	reading comprehension	N/A	research paper	N/A	N/A	N/A	university	course	
4	phonics	Sounds Write programme	resource	N/A	company	N/A	N/A	digital media	email
5	reading proficiency	Scarborough's Reading Rope	theory	Holis Scarborough, Haskins Laboratories	other research organisation	Sounds Write	company	training	

Excel was used to record and log all teachers' mentions of research, this work was carried out based on the transcriptions. All the transcriptions were read through and all 'mentions' were recorded into an Excel file, see above a modified example from the spreadsheet of recorded mentions.

The second tool to "organise" information we used was **N-Vivo**. NVivo is CAQDAS, computer-assisted qualitative data analysis software, widely used for qualitative and mixed methods research analysis as it is intended for the analysis of unstructured data of various types, see the interface below.



NVivo was used for thematic analysis and the procedure is explained below.

Use of NVivo had various affordances, including that of enabling researchers to **code simultaneously**, record observations and questions as memos and easily explore coded data. But it also fragmented the data, divorcing extracts from the spaces in which they were generated, removing visual and physical cues, temporarily deleting objects, other participants and often obscuring the medium through which the exchanges had been recorded. Our collaborative, reflexive approach helped to maintain our awareness of these issues ([see below for more details](#)).

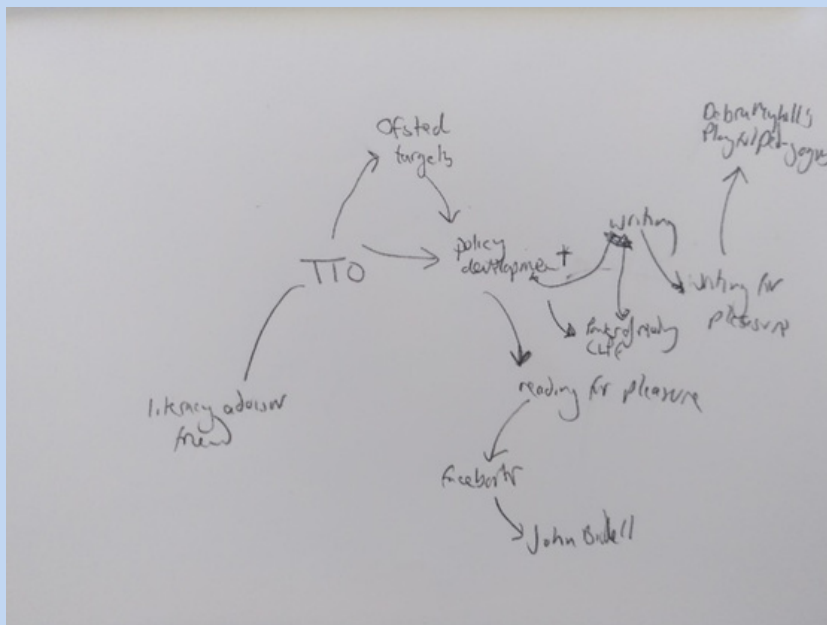
Approaches

We developed an iterative, interpretive approach to analysis, weaving together insights from **storying**, **thematic analysis** and a detailed analysis of **teachers' mentions of research**. These provided us with different entry points to the data analysis.

Storying

Storying generated **vignettes and mappings of individual teacher's encounters** with research. It enabled us to draw out complex movements of research, to tease out how teachers felt positioned in relation to research and to trace connections and discontinuities.

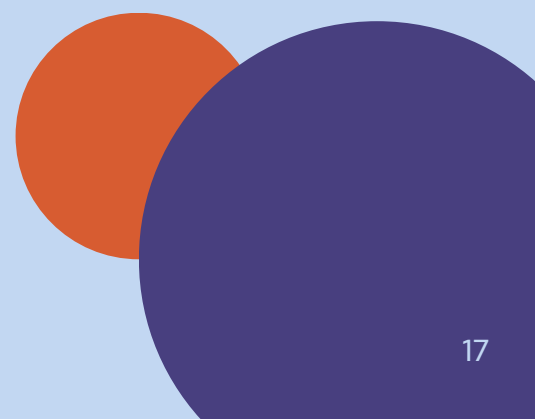
It felt important that the stories, rather than being held to represent singular truths, drew readers' attention to multiple interpretations and possibilities (McCormack 2004).



Our approach to storying was informed by Gill's background in narrative research developing interpretive stories from interview data and Cathy's development of 'stacking stories', an approach juxtaposing multiple tellings of an event from different perspectives with the intention of provoking "openings, contradictions and loose ends" (Burnett & Merchant 2020, p. 80).

We began storying by rereading all the transcripts, while asking the question 'As we read, with research mobilities/immobilities in mind, what strikes us?' We wrote short 'stories' drawing out movements and noting questions that surfaced as we attempted to develop an understanding of teachers' encounters with research. We swapped stories, commenting on each other's, developing existing threads and exploring new ones. This **iterative process** enabled us to raise new questions and begin to map the complexities of movements of primary literacy research to, from and around teachers. The stories drew on Burnett and Merchant's stacking stories (Burnett & Merchant 2020, p. 80) method using different stories of events to represent multiplicity, to open up possibilities, with stories "providing a different take but each also perhaps troubling the last".

As we wrote, we attempted to produce visual representations of research movements. The sketches or mappings that we produced were simplifications that helped us to begin to explore patterns. The mapping, see an example on the previous page, explores movements from an extract of an interview with one teacher, the arrows showing the work that the teacher was doing to seek out research.



Thematic Analysis

Thematic analysis enabled us to look across the logs, transcripts and images from the online meetings, examining fragments in depth, assigning codes, reflecting within and across codes to identify patterns in the data. Our analysis, analytic memos, reflections and regular conversations between ourselves and with the wider project team enabled us to develop understandings and draw out our findings. ‘Thematic analysis’ covers multiple approaches to exploring, analysing and interpreting patterns in qualitative data, using coding processes to develop themes (Braun & Clarke 2021). Thematic analysis has been described as “theoretically flexible” rather than having an underlying prescribed theory (Braun & Clarke 2021, p. 338); however, a reflexive stance was also important.

Coding

Coding is an interpretive process of “noticing potentially relevant meaning” (Braun & Clarke 2022, p. 236) and **labelling** it with a code. In reflexive thematic analysis, coding is organic and flexible, deepening as it progresses, unlike coding reliability approaches that tend to echo scientific methods with themes often developed prior to full analysis, guided by theory and research questions (Braun & Clarke 2022).

Prior to embarking on coding, we increased our familiarisation with the data through a process of reading and re-reading transcripts, checking against recordings and making notes. Initially, each segment of the transcript raised questions about the wider circumstances, about possible connections, what was missing and about how to code. We worked on these questions in our **analytic memos** and through discussions. Our discussions enabled us to probe how key terms were understood and used by participants and other researchers. Analytic memos (Saldaña 2021) took the form of journal entries, notes, reflections and questions throughout the analytic process.

'Research' itself emerged as a very slippery term, we mused on this, for example, in the following memo:

For 'teacher school led research' I'm coding where teachers talk about doing some research themselves, including where this sounds as tho' what they are doing is further reading or exploration e.g. T43 in Int 9 says

"What I tend to do is leave the email unread until I can read it, until I've got some time to read it, then I know that will usually follow through to me clicking on something or following it up with some research myself.

As we developed our coding, we decided against codebook approaches in favour of a more open exploratory approach. However, we noted that coding topics, actors and evidence types might entail a more descriptive approach, with reflexive thematic analysis likely to be more appropriate for other codes.

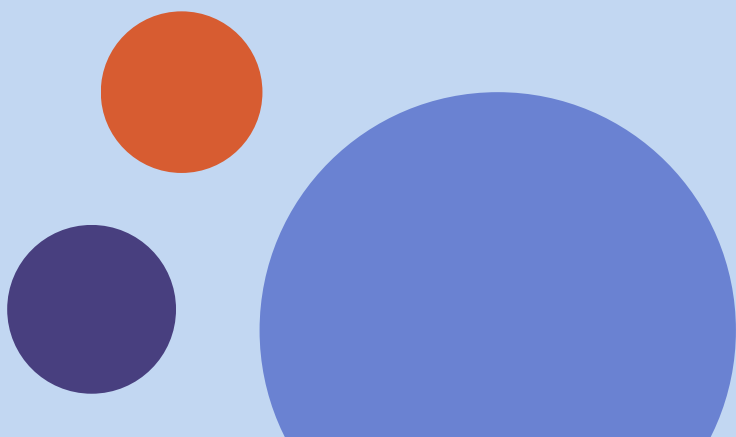
These two coding approaches were developed through discussion. As we continued our familiarisation with the data through the process of coding, we noticed additional fragments of data that were potentially relevant, resulting in new codes. Examples were 'absences' (any topics, encounters, or actors that were mentioned or implied as missing, not there or hidden) and 'multiple mentions' (what happens when the same topic/research etc comes up from different sources).

The descriptive coding, as in the excel sheet as explained above and further below, also informed the early development of codes. For example, we had initially proposed a code *actors* with the description 'Individuals, organisations, texts, policy, schemes, social media, events, places and spaces, commercial'. After discussion, we split the code and coded different actors such as *brokers* (for individuals and organisations), *channels* (for social media, blogs), and *resources* (for schemes) separately.

This raised questions about what these different approaches we used in parallel did - how we might bring them together - as well as how the thematic analysis interacts with the stories. We ran these approaches in parallel, with themes actively created by us. We began by coding the data, using NVivo.

The monthly team meetings and other interactions across the different strands of the project informed coding. For example, as we coded, the foci for the cases were emerging and we added codes to help trace these in the teacher data, for example, *Ofsted English review* or *Reflecting Realities*. Codes also developed as a result of conversations with the wider project team, for example, we talked about the difficulties with the codes *channel* and *broker* and how we distinguished between them, resulting in a decision to merge them.

The simultaneous work on the spreadsheet of 'mentions' in Excel prompted changes in terminology and categorisations, for example, we refined codes further by, for example, changing *text or scheme* to *resource*.



An example was a recurrent thread that seemed significant: “**an assertion that research is behind or at the bottom of an initiative but not visible to participants**” (analytic memo) leading to us later adding a code: *research behind it assumed*. Despite this iterative, collaborative process, coding remained challenging as this extract from a researcher’s journal shows:

I’m struggling to clearly differentiate between ‘interactions with research’ and ‘movements’ – the descriptions seem entangled and I’m frequently dual coding

Once we had worked through the data several times, revising and finalising codes and feeling confident that the code labels and descriptions represented the diversity of teachers’ encounters and communicated our interest in those features, we started to think about connections between codes and developing themes.

Generating themes

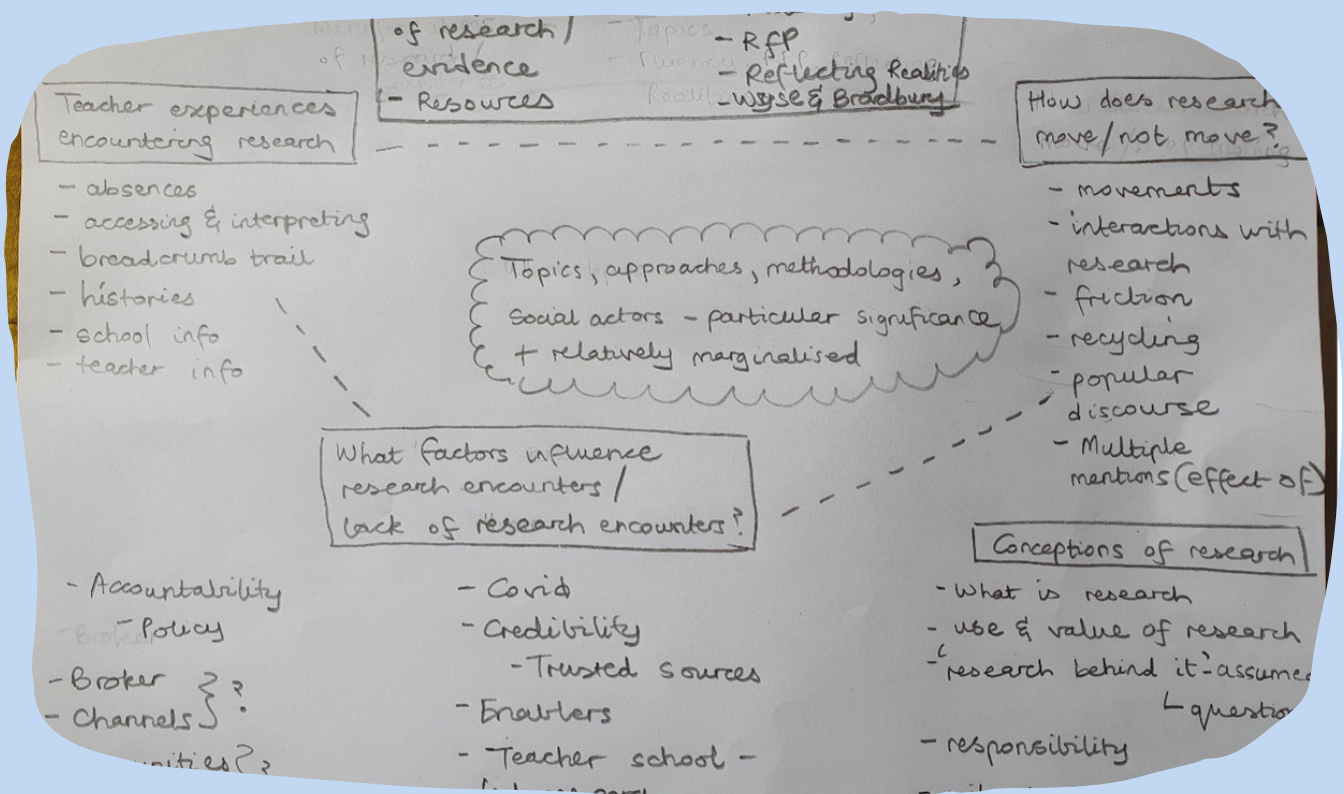
Initial themes were developed from codes by exploring areas of similar meaning across codes and clustering these potentially connected codes. We began by looking at the list of all codes and their descriptions alongside the research questions. From this, we developed **visual maps** constructing these individually before discussion. Following discussion, we reviewed our initial themes/mapping. Initially we were guided by Braun and Clarke’s (2022) questions to support theme development and revision.

First, we considered whether the patterns we had found had “an identifiable central organising concept” (Braun & Clarke 2022, p. 98). This helped us clarify what the theme was about and identify which codes contributed to it. An initial theme *teachers’ research relationships* developed into *research spaces, conversations & relationships* as we discussed decentering teachers, ensuring that a range of actors were included. Another initial theme, *perceived local relevance* included a number of codes, several of which (*communities, enablers* and *critical stance*), were later linked to the revised *research space, conversations and relationships* theme.

One code, *credibility*, was promoted to a theme as it became clearer through our analysis and discussions that this was a distinctive feature of teachers’ encounters with primary literacy research. Where data extracts were assigned to multiple codes, we revisited the extracts to attempt to determine the most salient. The following extract shows the multiplicity – we felt it fits better *research spaces, conversations and relationships* but it is also coded as *communities*.

An interesting thing was that last week I went on an HMI English lead, just like a short half an hour on what to expect during inspection and a number of teachers there all said we really miss the idea of meeting as a consortium, so meeting together with local schools and I thought back to reflecting what we had said last time, that it is a thing of the old days almost isn’t it, and yet really that is the perfect forum to discuss bits of research that have been useful and I think that is lacking a little bit. We were all talking about, the HMI said where do you get your information from, and various schools said ‘Oh EEF documents and we research bits and pieces’ and it became apparent that really, if you had that sort of forum of a consortium that I think you would get a lot of information out of that as well. (T43)

Initial theme development was anything but straightforward. For example, one of us explored using a mind mapping tool in Nvivo to develop a map of themes but found them “frustrating as they restrict the relationships and structure of the map” (field notes). As we had found in other stages of analysis, what was most valuable was sharing our work in progress with each other, articulating the rationale for grouping particular codes together, posing and responding to questions before a further round of review and revision. See below examples of the initial maps of codes developed by Gill and Pari.



We continued to work iteratively, generating tentative themes, reviewing and revising as we worked, whether there was sufficient meaningful and rich data to support the theme and whether the theme was important in relation to the data and the research questions (Braun & Clarke 2022, p. 99). We used Braun and Clarke's questions to support this review, for example, considering whether we could identify theme boundaries. Here we worked with our initial themes, through several iterations, to develop theme descriptions. For example, for the **credibility** theme we grouped the codes *trusted sources, credibility, popular discourse, channels, manifestations, evidence type, breadcrumb trail* and *multiple mentions*, developing an initial description:

This theme focuses on credibility and how some sources became seen as more or less trusted or trustworthy. It explores how credibility appears to have been achieved, or not. We identify several markers of credibility and the way these overlap. We begin by exploring the role of multiple mentions in conferring credibility, before drawing out examples of signifiers including professional and relational credibility and lineage. Other markers include ratification through publication, resonance with popular discourse and official status.

One of Braun and Clarke's (2022) questions focusses on boundaries, asking researchers whether the boundaries of a theme are identifiable, with clarity about what it includes and excludes. This raised questions for us - we found that **we both could and could not identify boundaries of themes**. This was evident at the level of individual data extracts and at the level of the overarching themes. This was particularly difficult where data extracts clearly fit more than one theme - and this seemed a frequent occurrence. We noted that a feature of the data and the themes was "**complex messiness, unpredictability, uncertainty, always open to review**" (analytic memo).

Our final themes appeared more like perspectives – so rather than there being firm boundaries, what was held (temporarily?) together in a theme was at the same time in interaction with other themes and shifted a little as we continued to engage in analysis, with theory and with each other. This raised questions about how we conceptualised themes and how this conceptualisation was entangled with “typical” representations of themes, often as lists or in tables, sometimes as a simple network showing relationships. We questioned whether these representations worked for our project, this data set, these research questions. Would perspectives or fluid understandings been more appropriate than themes? **Might a mesh, a map, or a kaleidoscope help us represent the patterns and relationships more clearly?** Or possibly webs, entwined – interrupted to exemplify particular perspectives, providing glimpses into teachers’ stories?

Visualisations

We used a variety of visualisations in this part of the project: **mappings as part of data analysis, visualisations in data generation with teachers and artistic responses as part of our data presentation.**

As noted above, an early stage of data analysis involved us making sketches to map how we saw research moving in the teachers’ accounts. For this, we worked with the interview data, initially focussing on one research encounter, then expanding our gaze to consider how this encounter related to others that the teacher talked about. We shared these mappings, along with our stories of these encounters, the questions and discussion that this prompted helping us to see both what the teachers talked about, **what was present, and what was absent.** We also used mappings in the thematic analysis, reviewing codes and beginning to generate themes.

In adapting our approach to the work with teachers we asked them to create a visual representation (using online tools or paper drawing) of how they saw research moving to and from them, one that looked across their encounters with research. Initially, we used a padlet, an online board that enabled participants to post comments and images and move their posts around. This was productive in the online group space, enabling participants to see and interact with each other's posts. Later, particularly where there was only one teacher, pen and paper drawings were more often used. Teachers shared their visualisations, talked through the representation, and responded to comments and questions. This approach was influenced by our initial analysis of the interview data from the first phase, described above.



Our artistic responses were produced by Lo Tierney
(www.instagram.com/lotierneyarttt)

Another way that we used visualisations in this part of the project was in the artist's responses to the teacher data. Our intention here was to work with an artist to attempt to represent the complexity of teachers' encounters with research and the way that these were entangled in their work and personal lives. We have written about our collaboration with **Lo Tierney** [in this blog](#) where you can also see the images produced.



The teacher up high on a ladder, selecting a book might be seen as reflection of the challenges teachers face in accessing research, the lengths (or in this case heights) they have to go to. They seem precarious, yet unconcerned, as if habituated to this precarity. I've just been working with the movements data again, thinking about chain reactions and this image represents a moment on one chain reaction described in the data fragment that acted as impetus for the drawing: the teacher has been following the Reading for Pleasure project, exploring teachers' mini research projects that have been uploaded to the project website, and as a result is planning an event for parents to support them to encourage children to read more widely.

The technologies used in the production of an image influence its form and contribute to the effect it has on viewers (Rose 2016). For example, Padlet enables users to search online for images, with search engine algorithms determining results to select from, or to upload their own images. NVivo allowed us to develop a map of themes, although as we have noted, the design of the mapping tool constrained possibilities. Our pen and paper mappings as part of the analysis were dependant on our skills and restricted into two dimensions by the medium.

In addition to the technologies used, compositional and social modalities are also important (Rose 2016). In engaging critically with visualisations we might consider for example, how compositional and social aspects are affected by where and how the image circulates and how it is viewed. We might ask what the maker or the research intended, and what the images do.



Teachers' 'mentions' of research

A third approach to analysing the teacher data involved capturing what teachers mentioned in relation to their encounters with literacy research and how they encountered this. We refer to these items as '**mentions**' because, while some items were described at length, many were brief passing references made as part of generalised statements about a range of reference points, contexts and sources of information. Our approach was to make an entry on the spreadsheet for all mentions that just might relate to literacy research, excluding only those where participants explicitly said they were referring to other kinds of resources or information.

This allowed us to gain insights into the range of research that they encountered and the individuals and organisations that mediated this research. We were then able to conduct follow up investigations to explore the kinds of sources teachers were accessing and the relationship to underpinning research.

We recorded all references they made to identifiable research sources and noted the range and relative prevalence of topics, research producers/sources, research brokers, and research mediating channels.

Mentions were logged on an Excel spreadsheet noting **topic, source, source type, origins** and any **intermediaries**. *Intermediaries* involved any individual, organisation or technology that played a role in the movement of research to, among or around teachers. The *topic* was divided into *general* and *specific*, for example, *reading* as a general topic and *reading comprehension* as specific. Topics could be very general, such as *verbal feedback*, or very specific *Rosenshine's principles*.

See the following extract, which has been recorded in line 4 in the Excel screenshot above, the detail reproduced below for convenience:

	Sounds Write programme	resource	Source of					
phonics specific topic	name	source type	N/A identifiable	company source type	N/A Broker name	N/A Broker type	digital media Channel type	email intermediaries

I trained on the Sounds Write Phonics Programme I have continued to read what they put out and I think that I mentioned previously that they have a podcast that I have listened to and actually just thinking back, I think that perhaps I might even have particularly sought out the first one because I was doing this literacy research and it all feeds around in a big circle, and I was looking at where I was connecting with literacy research and I thought oh why not, I've got that phonics podcast on my 'to listen' list, let's bump it up to the top and listen to that one and then I can write it down in the diary.

[00:11:07] But then because I'd listened to one I went 'Oh, this is actually really useful, this is helping me develop my practice', and then I went and listened to the next three as well as they came out. And so it's all sort of bumped around in a circle and I think that possibly I'm not sure if you do want to hear this, because you don't necessarily want to think that your research has influenced my behaviour but it definitely has and it's nudged me in to exploring new areas or not new areas, but going deeper to find things actually that I'm really interested in.

[...]

because I have done the Sounds Write training I'm on their mailing list, so when they started this podcast and released the first episode they sent out an email and I looked at the email and went gosh, who would ever listen to a podcast about phonics in their spare time?! And I clicked on the link anyway, so it was there on my podcast list and then I went 'Oh, let's give it a try'.

Sometimes the logging of mentions was straightforward but it frequently involved a degree of interpretation and entries were often incomplete. For example, a participant might mention that they use Twitter but provide no specific example of this, or they might mention a piece of research but not how they encountered it. All these incomplete references were logged in order to give a wide ranging map of the kinds of topics, research and mediators that featured in teachers' lives. The completeness of the Excel line depended on the context of the provided information, which oftentimes was vague or very difficult to pinpoint and consequently many cells remained empty.

In categorising mentions we drew on our own professional experience combined, where necessary, with internet searches to identify origins and/or the nature of organisations and individuals involved in producing or brokering research. This process was far from straightforward. For example, in relation to organisations and individuals, the category of *charity* included organisations with charitable status but with very different purposes. They included subject associations like UKLA, professional bodies like *Chartered College of Teaching*, as well as public facing charities such as *National Literacy Trust*.

Moreover, individuals often have multiple roles. Teachers and academics may also be consultants and many have dual roles within commercial, charitable, and/or educational organisations. And individuals and corporate websites are not always transparent about their aims or role. Several, for example, present as charitable bodies but are essentially commercial enterprises and many leverage previous experience as an academic and/or teacher for credibility in their bios on commercial websites.

This systematic noting of mentions was then subject to a process of consolidation to avoid repetitions – i.e. where a participant had referred to the same source on multiple occasions. For example, in the first phase of the project, eleven teachers participated in a sequence of activities that involved an initial workshop, a period of lifelogging, a workshop to reflect on what they had logged, a further period of lifelogging and a final interview or focus group to reflect on this second phase and to think across their experience.

Consequently, on occasions, the same item registered in more than one data source and, where the information was the same. these were combined into a single entry. Repeated mentions, however, were retained if the information within them was distinct in some way. For example, a participant might mention the *Open University Reading for Pleasure* website on one occasion and the *Open University Reading for Pleasure* conference on another - these were logged separately. Though initially, we were hoping to capture some quantified trends, the frequent uncertainty in categorisation made quantifying extremely difficult, if not impossible, and though we made several attempts, some of which were presented at conferences, in our final publications, we decided against them as we felt they would distort the complexity.

Given the nature of this data, it is important to note here that this logging of mentions should not be regarded as definitive or representative, either of these teachers' encounters with research or of the wider teacher body. It is quite possible that these teachers encountered research that they did not mention to us and that other teachers would have encountered very different sources. Instead, the spreadsheet provides a rich snapshot of literacy sources, origins and mediators known by a group of teachers at a specific point in time, a group of teachers who we might reasonably conclude are particularly engaged with research and/or literacy.

Perhaps most importantly, this spreadsheet acted as a portal to a vast range of research sources, sites and intermediaries that could be investigated further, for example through tracing the origins of texts, considering the relationship between content and format/presentation, and comparing appearances of research findings in multiple formats.

Investigating Research Mobilities: Methods Combined

The different approaches to **data generation** (workshops, focus groups, interviews and lifelogging) and **analysis** in combination enabled us to investigate teachers' encounters with research in their daily lives, within and across the increasingly blurred boundaries between work and home, as they sought out and came across research on variety of channels, propelled by various mediators. Lifelogging together with conversations with teachers in interviews and focus groups, helped us to see what research teachers encountered, to explore how their encounters were situated, how they connected to other aspects of teachers' lives and elicit participants' reflections on these. The artists' interpretations of the data, juxtaposed with extracts from transcripts, acted as striking visual reminders that multiple interpretations of the data were possible.



Corpus Linguistics

Introduction

Anna and Julia

Corpus linguistics is an empirical approach to studying language. It uses large collections of real-world text data to uncover **linguistic patterns and trends**. A 'corpus' (or 'corpora' in plural) is a large and structured set of texts in digital format; these can be **written, spoken**, or even **multimodal**. Corpora would often also have some layer of **annotation**, for example, part-of-speech tagging that helps to look for specific types of words, such as verbs, or lemmatisation, which groups word forms together, for example *read, reads, reading* could be all found under lemma READ. Some of the key characteristics and applications of corpus linguistics include:

- **Data-Driven analysis** – the emphasis is on examining and describing actual language use in natural contexts.
- **Large collections of texts** – a corpus is a (usually large) systematically organized collection of texts that represents a language, or a specific part of a language, for example, casual conversation. A corpus can include anything from literary works, newspapers, and online texts to transcriptions of spoken language. Some of the existing corpora run to billions of words but much smaller corpora are also used; these are typically corpora representing a specialised language use, for example, a corpus of works of Charles Dickens is less than 4 million words.

- **Quantitative and qualitative methods of analysis** - both quantitative (statistical analysis and frequency counts) and qualitative (contextual analysis) methods are used. This dual approach allows for the investigation of linguistic patterns and trends in language use, as well as in-depth analysis of specific linguistic phenomena. So, for example, large general corpora show us that one of the most frequent nouns in English is the noun *time*. When we look closer at how the word is used, we uncover that the patterns around time include familiar expressions such as *time to* [do something], *time of* [e.g. *your life/(the) year/the day* etc], *time for* [something] and other very frequently used phrases.
- **Natural Language Processing (NLP)** - Corpus linguistics has a significant impact on computational linguistics and NLP, providing the data and insights necessary for developing and refining algorithms used in NLP applications such as machine translation, speech recognition, text-to-speech, and other language technologies.
- **Variety of applications** - apart from linguistics, corpus linguistics is used in various other fields, including literary studies, cultural studies, psychology, and sociology, to analyze how language reflects and influences social and cultural phenomena.

Corpus linguistics is used in investigating some of the many existing, often very large, corpora, including also other languages, rather than only English. Some of these corpora are of general nature and you can easily look up usage of words, including in specialised registers; other corpora are highly specialised, including, for example, spoken or historical language. For contemporary British English, for example, the new *British National Corpus 2014* is a valuable resource, or *COCA* for American English (for further resources please check, for example, [here](#)). An alternative approach, adopted in this project, is to create smaller, specialist corpora.

In the *Research Mobilities* project, corpus linguistics methods were primarily used to answer the research question ‘Which literacy research is circulating?’ We focused on two different types of corpora, both, in a way, serving as a snapshot, with no claims to representativeness, of public discourse around the themes of literacy. We created two corpora for this project: a newspaper texts corpus to see which topics the media ecosystem is interested in, and a corpus of social media interactions – less formal, more varied, more geographically spread, but also more specialised.

Data collection

We created and used three small-size specialized corpora that have been specifically collected for the project. We used three types of textual data:

- newspaper texts (media corpus);
- Twitter data
- the transcription data collected during the “teachers’ encounters”.

All three corpora are snapshots of different discourses around literacy topics. The “teachers’ encounters” corpus represents highly specialized professional spoken discourse, elicited in workshops and interviews, with agenda and, to some extent, themes set by our research team. Our newspaper corpus represents discourse for audiences with both specialized and broad interest in our topic, as set and guided by media professionals and, to some extent, also experts writing for newspaper outlets.

The Twitter corpus also represents specialized discourse both in terms of its content and format. Twitter (now X), in its format, being distinctive with its short messages with frequently linked multimedia content, was, at the time, perceived as an important social media outlet, complementing traditional newspapers in the contemporary “hybrid media system” (Chadwick 2017). Twitter was also understood, based on previous studies, as being prolifically used by teachers to share and exchange ideas and content (Guest 2017). Things have likely changed substantially after Elon Musk’s takeover of the platform, highlighting the volatility of social media spaces.

Media corpus



Newspaper collections are widely used in **corpus linguistics and discourse analysis** to explore how various ideologies are disseminated through the media's gatekeeping processes. Education is an area that commands the attention of a wide range of stakeholders including children, parents, and public figures, beyond just experts, reflecting its critical role in society and the significant impact that media discourse can have on shaping public perceptions and policy. The media gatekeeping ecosystem represents an intricate system through which information is curated, shaped, and shared with the public, underscoring the media's influence in determining which ideas and viewpoints reach a wider audience. The media (newspaper) corpus was therefore envisaged as **a proxy of public discourse**, i.e. including also more general, less specialised audience.

In our project, we collected the newspaper texts from the Nexis News database, which is part of the commercial LexisNexis platform, to which we had access through our university library subscription. The platform provides a considerable amount of resources among which is the News database. The database allows its users to construct sophisticated queries based on its internal structure: the user can specify, in addition to the search terms, for example, timeline, geographical region or a particular newspaper. While these features allow you to conveniently narrow down your results, they can be also in some cases restrictive.

One of the key decisions in data collection like this are the search terms: very specific terms or too many terms will considerably narrow down the results, more general terms, on the other hand, may produce substantial amount of “noise” (irrelevant data). Researchers also need to be aware of register differences, the words they use in their specialised jargon (e.g. *literacy*, *research* in our case), may have different meanings in more general registers, such as newspapers. Often, the best way to go is experimenting: for example, we started with a search that included terms *literacy*, *research* and *primary* – because we were interested in ‘research mobilities’ in ‘primary literacy’ context. However, this combination produced substantial amount of articles where the words *primary* and *research* referred to *primary research* (in all sorts of disciplines) rather than *primary literacy*.

We have also considered using keywords referring to specific literacy areas such as *reading* and *writing*; however, these being very frequent words in general language, the results produced huge amount of noise. In the final round of fine-tuning the keywords, we also decided to drop the word *research* as it not only produced noise in the data collection as mentioned, the word itself seemed limiting. *Research* is used in the newspapers in numerous ways and not necessarily only referring to 'research' as conceptualised in academia. Additional terms that are used in the sense relevant for our project include, for example, *study*, *project*, *article*, *survey*. We felt that limiting the keywords by adding *research* may actually leave out many relevant results. Due to time constraints in the data collection phase, we settled on terms *primary*, *literacy* and *education*. This, necessarily, of course, means, that our corpus covers to some extent appearances of literacy more generally; however, the affordances of corpus linguistics methods, are well equipped to distinguish between these cases. Our collection timeline was decided arbitrarily to cover 5-year period, starting on 01/01/2017 and going somewhat beyond the 5 years up to 09/05/2022 (the cut-off point was given by the date of the final collection). It is worth noting, that this period covers the 'lockdown' with its school closures.

As the project was concerned with research mobilities in England, we also narrowed the search to 'England and Wales' (it is not possible to have 'England' only in the Nexis), see the screenshot from the Nexis database below.

The screenshot shows the Nexis database search interface. The search query is 'primary AND literacy AND education'. The results are filtered to 'News' and 'England & Wales'. The search results are sorted by 'Relevance'. The first result is a news article titled 'Wandsworth pupils first ever to visit the Centre for Literacy in Primary Education's revamped library' from the Wandsworth Times, dated 20 Dec 2018. The article text is partially visible, mentioning 'Wandsworth pupils first ever to visit the Centre for Literacy in Primary Education's revamped library...' and 'Four local pupils from Wandsworth Prep School have been the first ever group of primary school children to visit the refurbished library at the Centre for Literacy in Primary Education. The CLPE (Centre for Literacy in Primary Education) has one... Wandsworth Prep Year 6 Reading Leaders to be the first pupils in the country to visit the Centre for Literacy in Primary Education's new library. It was an especially special visit for one pupil who was able to see an illustration drawn on the wall by Chris... about a range of books and it was great to hear that the children were taking back recommendations to share with others at school as well as adding to their personal reading lists.' The four children chosen to visit the Centre for Literacy in Primary... Education were the school's Reading Leaders. The school's Reading Leaders support the teachers who have the responsibility of English across the school. They are involved in the organisation of the library, organising various book events and carry out...

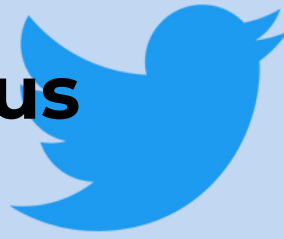
The collected data was then exported and manually checked to remove duplicates and irrelevant results; after this procedure we had a total of **426 newspaper articles (405,603 words)**. The articles were then “cleaned” to contain only the text of the articles and saved in .txt format under a unique identifier. We created an Excel sheet with accompanying metadata, where we recorded the date of publication, the newspaper the article was published in, the author, the article title and length in words. The metadata also contains information on whether the article included any graphical elements.

Media texts are not a monolithic resource, each newspaper is characterised by its unique characteristics in terms of intended audiences or political orientation. In the UK, for example, newspapers are often divided into tabloids and broadsheets. This distinction was less relevant for our final corpus as it included only very few tabloid articles (it can perhaps be tentatively argued that the topic of primary literacy seems to be of less interest for tabloid papers; however, it is also likely that it was our choice of the search terms that skewed our corpus towards broadsheets, it is perhaps the rather technical term *literacy* that produces this skew).

When working with newspaper data, some basic characteristics of newspaper writing are worth considering, for example, broadsheets tend to have longer articles than tabloids and regional papers, the vocabulary used also tends to be different. We have, therefore, divided our corpus into three subcorpora: national papers (157 articles, mostly broadsheets), regional papers (183 articles) and the *Times Educational Supplement* (86 articles), a specialized national newspaper popular with teachers and educationalists (please note, the *Times Educational Supplement*, *tes*, has become from 2022 an online publication only, and consequently we do not have any *tes* articles from 2022).

This division is important when considering frequencies; so while in terms of the number of articles included in the corpus, national newspapers represent 37% of the corpus and regional papers 43%, it is the other way round in relation to the number of words – national newspapers take up 31% and regional newspapers 27%. This is due to the article length. Perhaps unsurprisingly, the biggest part of the corpus in terms of the word count is actually made of articles published in *tes*. For any further analysis and interpretation of the findings, this is important to keep in mind as *tes* is a highly specialized publication.

Twitter corpus



Twitter (now X, it was Twitter at the time of the collection and therefore will be referred to as Twitter) was chosen at the time because it was perceived as a social media platform relevant to our topic and it had been well documented as used by teachers (Burnett et al. 2022). Twitter having changed to X after Elon Musk takeover has undergone significant changes and its suitability for this type of research is no longer as clear. Even though takeovers like this do not happen every day, it highlights the unstable nature of social media platforms more generally. Any research into social media provides mere snapshots that are bound to time, both of posting and data collection, that are constrained by underlying opaque algorithms. The following relates to our specific project but would be, in many ways to an extent, applicable to other types of social media research.

In order to collect Twitter data, we applied for Twitter Developer Account to obtain our Twitter API Key (this was free at the time of our collection, the conditions have changed shortly after that). Working in the ESRC Centre for Corpus Approaches to Social Science at Lancaster University, we were fortunate to access **Twitter Collector** tool (Joulain-Jay 2021), with a user friendly interface for Twitter data collection. After some time of experimenting with our search terms (see the discussion above), we settled on: *primary*, *school* and *literacy*. We collected three years of data from 01/01/2019 to 31/12/2022. It is worth noting, however, that it is not possible to collect all tweets that fulfill the above parameters, Twitter has its own (generally unknown) algorithms for filtering and providing results to researchers, so for example, the dataset we collected seems to be skewed towards accounts based in the UK.

The collected Twitter data can be exported in several formats. For those unfamiliar with the Twitter data structure, the figures at the next page illustrate the Twitter data structure and show what a tweet looks like at the platform and what it looks like exported in an .xml format.

On October 27, 2022, user @Eve_Morton posted a retweet, that is, she reposted a tweet by another user, in this case @whatSFSaid. The pictures below show the retweet as on the social media site, a screenshot of @Eve_Morton's profile information (the sender of the retweet) and a screenshot of @whatSFSaid's profile information (the sender of the original tweet). The bottom picture shows how this looks in .xml.



Eve Morton
@Eve_Morton17

Primary Literacy Adviser for 'Education North Tyneside'. Teacher Reading Group leader with OU/UKLA School governor. SCITT lecturer. NPQLL facilitator

Newcastle Upon Tyne, England Joined April 2013

669 Following 743 Followers

SF Said - away now, back in late Jan
@whatSFSaid

Author of Varjak Paw, The Outlaw Varjak Paw, Phoenix - and now TYGER! I also write about children's books & all things connected to them.

London sfsaid.com Joined November 2013

14.3K Following 33K Followers



```
<tweet id='1585552391780405249' createdAt='2022-10-27T08:42:02.000Z' language='en' authorId='1379685786' authorUsername='Eve_Morton17' authorName='Eve Morton' authorVerified='FALSE' authorDescription='Primary Literacy Adviser for 'Education North Tyneside'. Teacher Reading Group leader with OU/UKLA School governor. SCITT lecturer. All views my own.' authorLocation='Newcastle Upon Tyne, England' authorCreatedAt='2013-04-25T15:22:31.000Z' authorFollowersCount='691' authorFollowingCount='611' authorTweetCount='2297' authorListedCount='4' referencedTweetId='1585545307861000193' referencedTweetCreatedAt='2022-10-27T08:13:54.000Z' referencedTweetText='What&apos;s the most important thing in education? If a child leaves primary school with a genuine love of reading, that will have a bigger positive impact on their life chances than any &quot;measurable outcome&quot;. There is nothing more important than READING FOR PLEASURE. #literacy' referencedTweetRetweetCount='193' referencedTweetReplyCount='6' referencedTweetLikeCount='705' referencedTweetQuoteCount='6' referencedTweetAuthorUsername='whatSFSaid' referencedTweetAuthorName='SF Said' referencedTweetAuthorVerified='TRUE' referencedTweetAuthorDescription='Author of Varjak Paw, The Outlaw Varjak Paw, Phoenix - and now TYGER! I also write about children&apos;s books &amp; all things connected to them.' referencedTweetAuthorLocation='London' referencedTweetAuthorCreatedAt='2013-11-11T22:06:59.000Z' referencedTweetAuthorFollowersCount='28762' referencedTweetAuthorFollowingCount='13285' referencedTweetAuthorTweetCount='79661' referencedTweetAuthorListedCount='312' retweetCount='193' replyCount='0' likeCount='0' quoteCount='0' >RT @whatSFSaid: What&apos;s the most important thing in education? If a child leaves primary school with a genuine love of reading, that will ha...</tweet>
```

In the screenshot of the .xml file, <createdAt=> is followed by the date when the retweet was posted, <authorUsername=> tells us who is the initiator of the post, <authorName=> shows the name of the author, <authorDescription=> is the short description that Twitter users have, location, number of followers and other information available follows. Since this is a retweet, we also have information about the author of the original tweet, including the metrics - retweet, reply, like and quote counts that show us how “successful” the tweet was. The text of the original tweet is at the end (please note, the author descriptions, follower and metrics counts might be somewhat different because Twitter is a dynamic platform and the data collection point was at a different time than the screen captures).

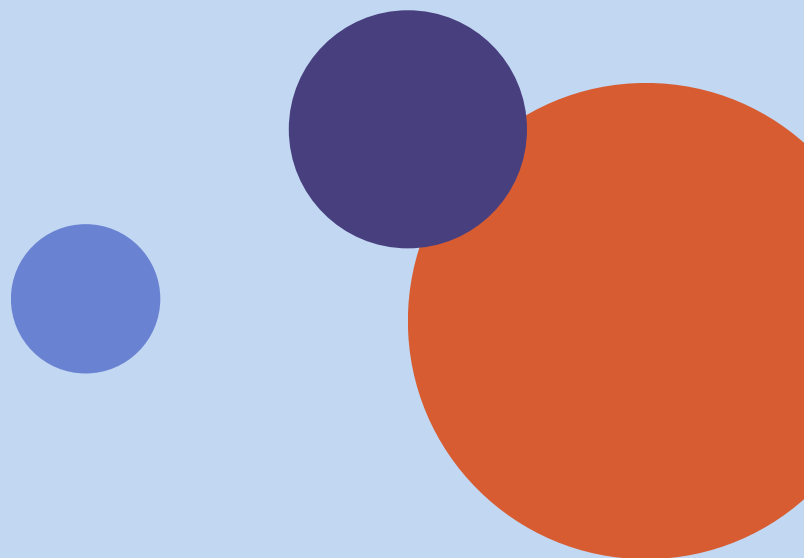
Another export format for the Twitter data included .csv files, which provides the same information in a tabular format (columns and rows); as we processed the data further in Python, the most convenient format for us was the .csv.

There are potentially many ways to explore social media data. We were primarily interested in what people talk about on Twitter. The most frequent types of interactions on Twitter are ‘likes’ and ‘retweets’ (as explained above), ‘replies’ and ‘quotes’ are much less frequent. All of these interactions, however, somehow react to the original ‘tweet’ and are thus, in one way or another, an echo, a repetition of the tweet. We have collected altogether **31,611 of these interactions** and to make things more manageable, we decided to focus only on tweets, which we have further divided into groups depending on the number of their retweets.

After some exploration, the cutting point for us was four retweets but, admittedly, this is a rather arbitrary cut-off point, which seemed to work for our dataset. To do this, we used a Python script to isolate the tweet texts and divided them into groups based on the number of their retweets. To reflect the individual tweets’ characteristics, we saved these tiny texts with headings containing the author name, the date the tweet was posted, number of retweets, replies, quotes and likes it has received. These were saved as .txt files so that we could work with them easily in other corpus linguistics software tools. The total number of tweets we had was 7186, amounting **271,784 words**. The file name looked like this: ‘X_2019-01-23_RT0_RP0_Q0_L4’ meaning X posted a tweet on 23 January 2019 with no retweets, no replies, no quotes and was liked 4 times. Out of the collected tweets, **only 14% of tweets were retweeted more than four times**.

Another way into the Twitter data set like this can be, for example, exploring user networks (for example with [Gephi](#)) or geographical spread (though location is not an obligatory information) and then perhaps plotting these on to a map (a very easy to use free online software is for example [Palladio](#) developed by the Stanford University).

We were interested, among other things, who the “influencers” in our specific discourse are (based on the number of followers and other metrics), who the discourse participants are more generally (exploring their ‘author descriptions’), and looking at themes being discussed through hashtags. Some basic coding skills are required to explore the data in this way, we have performed these explorations in Python. With the latest developments in the AI LLMs (Large Language Models such as ChatGPT), everyone will be able to explore this kind of data even without coding skills; though, very good understanding of your data to be able to check the accuracy of the LLM’s output will still be required.



Teachers' encounters corpus

The data for this corpus was collected in 2022 during the workshops, focus groups and interviews we held with the teachers participating in our project. The workshops, focus groups and interviews were recorded and then transcribed. The “teachers' encounters” corpus consists of 34 files of **transcribed recordings** amounting to around 277,000 words. Our transcriber received from us a set of transcribing conventions.

When working with transcriptions of spoken data, transcription conventions need to be decided in line with the research objectives. Transcription is an important methodological step, as the level of detail included limits (or opens up) the future analysis. Though transcription software exists, transcription can still be fairly lengthy process; so while, for example, a sociolinguistic study would likely need a very detailed transcription including pronunciation and intonation features, and possibly also such features as gestures and length of pauses, for our purpose, this level of detail was not needed.

We very loosely adapted the existing standards used in the transcription of spoken language, thus saving time and resources at this stage of the project. For discussion of transcriptions conventions see, for example, Love et al. (2017) and Collins and Hardie (2022), both publications are open access. See below a short extract from one of the conversations and how a simplified transcription may look:

TP3[??]: Ah we use that too. We had to do a massive fundraising effort to buy all the quality texts.

TP2: [@laughs.]

[??] : Sorry, I'm not sure what that is. Did you say CLPP?

TP1: Centre for Literacy in Primary Education – CLPE. They're brilliant. [=TP2 yeah]. They have all this great CPD and masses of resources. It's all really good stuff.

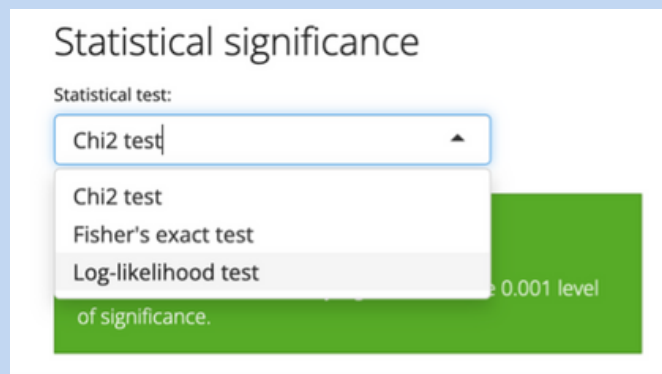
Analytical approaches

Corpus Linguistics way of working with textual data

Corpus linguistics analysis is essentially based on comparison and the usual starting point is a **comparison of frequencies**. We may want to compare frequencies of words in one dataset or across several datasets. For example, we have divided our media corpus into regional and national newspapers. The word *reading* occurs in the national newspapers subcorpus 519 times and the word *writing* occurs significantly less frequently, only 91 times. In the regional papers subcorpus, *reading* occurs 183 times and *writing* 103 times. These are called **raw frequencies**. While raw frequencies can be compared within one dataset, they **cannot be compared across datasets of different sizes**, as we are unable to determine whether the differences between these numbers are meaningful.

In order to be able to compare frequencies across corpora of different sizes, we need to **normalize the frequencies**. Many of the corpus linguistic software tools will do that for you but the formula is simple: the raw frequency count is divided by the number of words in the corpus, and then multiplied by whatever basis is chosen for norming, often it is 1 million. So, in this case, the calculations is as follows: 519 divided by 124,176 (number of words in our national papers subcorpus) multiplied by million - the normalised frequency is 4180, meaning that in a million words, *reading* will occur 4180 times, while *writing* will occur 829 times.

In comparison, in the regional papers, *reading* will occur 1669 times per million words. This is a starting point for further investigations. The frequency of *reading* in the two subcorpora seem different, whether the differences are statistically significant needs to be tested.



There are several tools available, one of the very easy ones to use is available here: <https://www.korpus.cz/calc/> (you may need to switch to the English interface version in top right hand corner). There are several options to choose from depending on what you want to compare, e.g. two words in one corpus, two words in two corpora. The statistical test includes a visualisation and gives you several choices of statistical tests. The significance level is set on the slider below the main table, see the screenshot below, which shows that the difference between the frequency of *reading* in the two newspapers subcorpora is statistically significant.

Calc: Corpus Calculator

- 1 word in 1 corpus
- 2 words in 1 corpus
- 2 words in 2 corpora
- 1 feature – many samples
- Many features – 1 sample
- Lexical richness (zTTR)
- N-gram correspondence
- About

Input

Word 1 frequency	Word 2 frequency
519	183
Corpus size A (tokens or words)	Corpus size B (tokens or words)
124176	109643

Significance level (α):

0.0001

0.01

0.05

If you need to compare frequencies of *two features in two different corpora*, you can use this module that is a generalised case of the module *2 words in 1 corpus* introduced above. It shares with it not only the means of evaluating the data (statistical significance tests and means of effect size calculation), but also the visualisation.

Example While the lemma *person* occurs only 571 times in corpus A compared to the 10189 occurrences in corpus B, taking into account the difference in size of the two corpora, we will come to realize that the relative frequency (ipm or ipw) is much higher for the former than for the latter. To make sure the difference is not accidental, we can add the confidence intervals to the relative frequencies. If the intervals do not overlap, we can assert (with a given margin of error) that the word is more frequent in the former of the two corpora.

Confidence intervals

Word 1: 4179.552 ipm (± 456.5972)
Word 2: 1669.053 ipm (± 300.4679)

Effect size

DIN d^2 : 42.925 (point estimate)

Risk ratio r^2 : 1.394 (point estimate; confidence interval: 1.315–1.477)
Ratio between relative frequency of Word 1 and Word 2 ranges from 1.315 to 1.477.

Statistical significance

Statistical test:

Chi2 test

P-value: 1.702e-28
Test statistic: 122.6047
The difference **is** statistically significant on the 0.01 level of significance.

Frequency comparisons can sometimes be very exciting and include a risk of “reading too much” into them. One should **never** draw conclusions based on frequencies only and **always** check the data against the larger context to see what is behind the numbers. Another consideration includes the nature of linguistic features that are being compared, essentially, they should be of a comparable nature. As a warning example of a linguistically extremely inadequate (and unfortunate) analysis drawing on words’ frequencies is the [analysis of Nicola Sturgeon’s resignation speech published in Daily Mail](#) (but also other media) on 16 February 2023, in which frequencies of pronouns (*I, me, my*) are compared to frequencies of proper nouns (*Scotland*). Not only pronouns are inherently much more frequent, considering the genre, i.e. resignation speech, they are to be expected.

The possibility that corpus linguistic tools offer, namely the flexibility to **switch easily between quantitative analysis and qualitative analytical probes** is, arguably, one of the biggest strengths of corpus linguistic methodology. One of the core tools in the qualitative analysis is the **concordance**.

Concordance

Concordance shows all the occurrences of the word we are examining (sometimes called ‘hits’) in a so called **KWIC** (key word in context) format: our search word is in the middle of the screen surrounded by a bit of context on both sides, see the screenshot at the next page. The concordance shows the first 16 lines of the concordance of *reading* in our Media corpus (screenshot taken from [AntConc](#) tool, see below)



Buxtorf Concordance Basel, 1632

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 1863 Page Size 100 hits 1 to 100 of 1863 hits

	File	Left Context	Hit	Right Context
1	180216DM.txt	blems. We now know that our efforts to improve the	reading	and writing skills of children from the poorest communities
2	200117TS.txt	wanted to help them in their mission to improve the	reading	and writing skills of children all over the country.
3	170422NP.txt	yd n Road, explaining to pupils how developing their	reading	and writing skills can lead them into all sorts
4	180622TNE.txt	' has hailed impressive improvements in youngsters'	reading	and writing skills five years on since the launch
5	190430BTA.txt	eatative ways parents can help develop their children's	reading	and writing skills at home. Children's author James
6	191213NP.txt	ar-olds in Nottingham (32 percent) did not have the	reading	and writing skills expected for their age by the
7	210326TTES.txt	. Also, while many teachers feel confident delivering	reading	and writing skills as part of their teaching, they
8	211105TTES.txt	pupils transitioning up to secondary school with poor	reading	and writing skills is a problem that has long
9	180622TNE.txt	o succeed." Middlesbrough Mayor Dave Budd added: "	Reading	and writing skills are one of the cornerstones of
10	180928STS.txt	ages – and our latest research shows that the joys of	reading	and writing can be hugely beneficial. "Not only does
11	180929SA.txt	ages – and our latest research shows that the joys of	reading	and writing can be hugely beneficial. "Not only does
12	190628TTES.txt	k. But success at interceptive timing didn't influence	reading	and writing development, so why the maths connection? "Some
13	210827TTES.txt	ers must be clear about the metacognitive aspects of	reading	and writing development, including deeper understanding of phonological
14	180928STS.txt	can be hugely beneficial. "Not only does a love of	reading	and writing enable children to flourish at school, but
15	180929SA.txt	can be hugely beneficial. "Not only does a love of	reading	and writing enable children to flourish at school, but
16	210707SA.txt	to just enjoy a bit of Beano silliness, creativity and	reading	and writing for the fun of it."

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

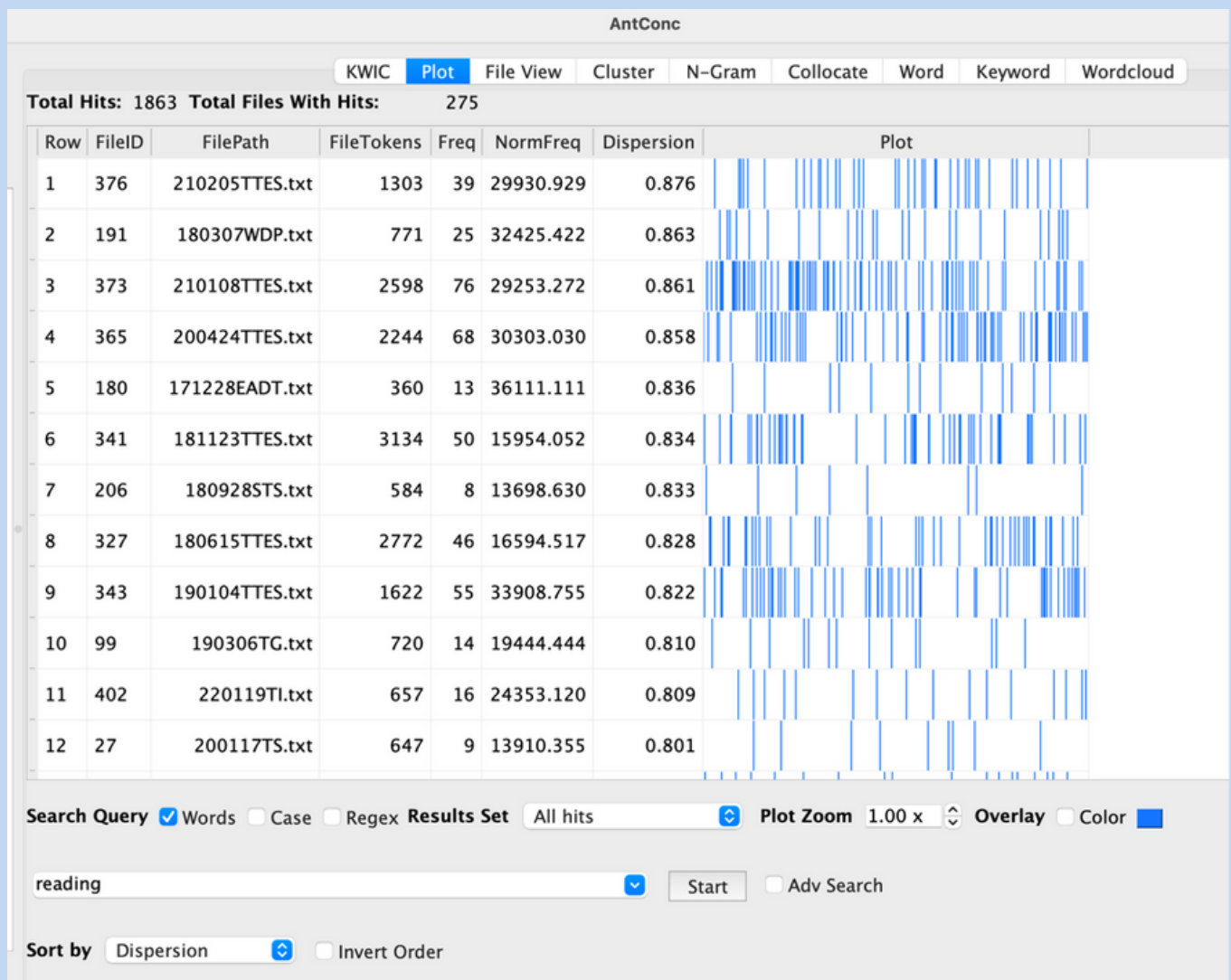
reading Start Adv Search

Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq

The concordance is **alphabetically sorted on the right hand side**, which allows for pattern identification, as in here, you can immediately notice the repetition of *reading and writing skills* and *reading and writing development*. You may also notice that some lines are repeated (lines 10 and 11 and lines 14 and 15). When examining the source texts (the first column), both of these come from regional papers where there is a practice of reprinting articles with only minor edits, or no edits at all (the analyst may then want to decide how to count cases like these).

In the concordance display, the user can specify a number of parameters, for example, **the size of the context or sorting parameters**. Most corpus linguistic software tools also allow to examine the full (or extended) context (here it would be in another screen, 'File view' at the top menu).

Another important thing to consider (particularly in the larger corpora) is the **distribution of the search word across the corpus**. Does it occur many times in just a handful of articles or does it occur consistently in many articles? The word *reading* may seem frequent in our Media corpus, but does it occur in the majority of the articles? We will find out that it occurs in 275 texts (out of 426) and some of these texts may mention *reading* only once, while in others, *reading* may be the main topic. To understand the distribution, most corpus linguistic software tools allows to create **plots**, see picture below. The screenshot shows how *reading* is dispersed in the individual texts in the corpus. Each vertical blue line indicates one particular occurrence of *reading* (and be clicked for extended context).



Collocation

While the concordance lines allow for a detailed analysis of the context and help to identify repeated patterns, one may ask what about if there are lots of them as in our case of *reading*. How do we analyse thousands of examples? One of the most common approaches in these cases in corpus linguistics is to generate a list of **collocations**.

Collocation, as a linguistic phenomenon, captures **a habitual co-occurrence of words**. Think of words you would, for example, expect to find in the context of *reading* – these could include *book* or *writing* as we have seen above. Corpus linguistics has come with a way to calculate the frequency of these co-occurrences based on the corpus you are looking at. Collocation is calculated as two words co-occurring within the proximity of each other **statistically significantly more frequently than by chance**.

There are several statistical measures that are used and every of these measures will produce slightly different sets of **collocates**. These collocate lists are not be taken as given but as suggestions for further consideration and analysis; however, the analyst should not just “cherry-pick” what suits them but needs to present explainable reasons why they discarded words from the collocation list. For example, common practice is to discard the so called function words such as *a, the, is, for...* and look at lexical words only. Having said that, sometimes it is the function words that can lead to a discovery of interesting patterns. Another parameter to consider when generating collocations is a **span**, that is how many words to the left and to the right you want the programme to look for. This may need some experimenting but as a very general rule, the span of five words to the left and right is a good span for English to start with. See the list of collocates of *reading* based on Likelihood measure on the next page.

AntConc

KWIC Plot File View Cluster N-Gram **Collocate** Word

Collocate Types 67 Collocate Tokens 3859 Page Size 100 hits 1 to 67 of 67 hits

	Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect
1	comprehension	1	82	10	72	22	275.100	3.712
2	love	2	89	87	2	65	269.292	3.461
3	pleasure	3	73	4	69	44	258.078	3.848
4	age	4	110	12	98	43	226.737	2.687
5	writing	5	91	7	84	56	150.266	2.335
6	enjoy	6	48	47	1	30	133.342	3.269
7	ages	7	37	4	33	13	122.715	3.685
8	enjoyment	8	35	15	20	25	111.342	3.582
9	motivation	9	29	4	25	3	89.269	3.503
10	skills	10	91	13	78	61	87.966	1.684
11	habits	11	23	0	23	18	76.450	3.691
12	aloud	12	20	0	20	13	68.696	3.775
13	expected	13	40	31	9	31	67.357	2.365
14	challenge	14	37	7	30	18	58.334	2.270
15	summer	15	33	24	9	13	58.036	2.431
16	children	16	264	159	105	127	57.891	0.726
17	books	17	99	32	67	61	55.894	1.232
18	acquisition	18	14	1	13	8	55.605	4.173
19	strategies	19	26	7	19	9	53.710	2.694

Search Query Words Case Regex Window Span From 5L To 5R Min. Freq 1

reading Start Adv Search

Sort by Likelihood Invert Order

The collocates can be explored further to see them in their context. See the concordance view of the collocate *love* (2nd collocate of *reading* as above).

File	Left Context	Hit	Right Context
13 190722TP...	has really helped motivate the children to read more. "A	love	of reading is something that can last a lifetime.
14 180928ST...	and writing can be hugely beneficial. "Not only does a	love	of reading and writing enable children to flourish at
15 180928ST...	e challenges, spending time with your child and encouraging a	love	of reading and writing is a good place to
16 180929SA....	and writing can be hugely beneficial. "Not only does a	love	of reading and writing enable children to flourish at
17 181025WS...	pleasure. Miss Miles added: "At Milton Park we encourage a	love	of reading and books from the minute they start
18 210415TT...	hopeless for the least well-off." Instead of encouraging a	love	of reading and writing, she says, schools have focused
19 210723mi....	selected in his book club which aims to foster a	love	of reading in disadvantaged children. Fifty thousand copies of
20 181229SA....	literacy Toolkit containing 50 books and resources to inspire a	love	of reading in them. The toolkits were funded after
21 201215EG...	wonderland as part of a determined drive to encourage a	love	of reading in pupils. Every spare area has been
22 220122TN...	brough Reads and share our mission of encouraging a lifelong	love	of reading in children, get in touch by visiting
23 190104TT...	a chance at a prize), than to instil a genuine	love	of reading in a six-year-old who is
24 181025WS...	six, with the hope that all children will have a	love	of reading for life. "Teachers choose high quality books
25 211002TN...	for a Book initiative aims to help foster an early	love	of reading for children in a fun manner that
26 220226CT...	between the two charities to help inspire and nurture a	love	of reading for over 130,000 primary school children who will

Keywords



Another way into the corpus data is through **keywords**. This method allows for a bottom-up exploration of the data – when you do not have a list of words you want to examine in the corpus, but rather want to explore what is in the corpus. A keyword (this is a different ‘keyword’ than the above mentioned in the KWIC) is a word that **occurs statistically frequently more often in your data than in a reference corpus**. Again, similarly as with the collocations, various statistical approaches exist to calculate the keywords (KWs) and there is no real consensus which is the “best”.

The ‘keyword’ method rests on the assumption that some words are more “typical” for each corpus. Mike Scott talks about KWs as revealing the “aboutness” of the text (e.g. Scott 2009, 2010; Scott & Tribble 2006). If function words occur on the list, they may be possible style indicators. For this reason, it is important to carefully consider the choice of the **reference corpus**, that is the corpus you are comparing to. This will largely depend on your research question. So, for example, if you are a literary scientist and you are interested in Victorian literature, you may want to explore the linguistic differences between Charles Dickens and Jane Austen. One possible way of doing this is through KWs and there would be two main approaches. You would either compare Dickens and Austen directly, using one as your target corpus and the other as your reference corpus and then swap. The other approach includes creating a third corpus of Victorian literature that does not contain Dickens nor Austen and compare both Dickens and Austen against this corpus.

One of the KWs comparisons we have done in the *Research Mobilities* project was comparing our Media corpus against a large newspaper corpus (published between 2007 and 2020 and containing about 20 million words), which is part of the BNC2014 (Brezina, Hawtie & McEnery 2021). The table on the next page shows the top results as generated by LancsBox and exported to Numbers.

Term	Focus rel. freq. (Literacy media corpus)	Reference rel. freq. (BNC2014: newspapers)	Simple maths	Log likelihood	% difference	Log ratio
literacy	2,640.32	4.57	26.2	7,708.31	57,642.10	9.17
reading	4,546.94	97.89	23.48	9,180.46	4,544.80	5.54
teachers	3,136.94	53.3	21.12	6,731.40	5,785.67	5.88
pupils	3,359.05	70.85	20.25	6,819.97	4,641.01	5.57
schools	4,706.66	146.23	19.52	8,510.67	3,118.77	5.01
primary	2,847.45	70.65	17.27	5,522.28	3,930.13	5.33
it's	1,424.97	0	15.25	4,506.96	Infinity	47.02
children's	1,312.67	0.05	14.12	4,137.27	2,669,680.56	14.7
education	3,196.83	138.75	13.81	5,126.41	2,204.00	4.53
learning	2,028.90	64.07	12.98	3,649.03	3,066.91	4.99
school	6,343.76	406.81	12.71	8,591.73	1,459.37	3.96
books	2,293.43	93.27	12.38	3,770.81	2,358.88	4.62
children	7,289.58	499.55	12.33	9,558.73	1,359.24	3.87
teaching	1,622.12	43.32	12.02	3,077.46	3,644.79	5.23
curriculum	1,033.17	7.87	10.51	2,594.64	13,033.19	7.04

N-grams

N-grams are another commonly used technique. **N-gram** is a string of words that is repeated across the text(s). The N stands for the number of words, i.e. the length of the string, a bigram consists of two words, while 5-grams of five words. The bigger the N, the fewer will be extracted from the text. N-grams are also sometimes referred to as **clusters** and **lexical bundles**. N-grams can be extracted in several ways, depending on the analysis aim. A bottom-up approach is to set the length (the N) and let the software generate all the N-grams of that particular length, with large datasets, we may want to consider a cut-off point for the repetitions, e.g. all N-grams that occur at least five times. The picture on the next page shows the first ten 5-grams from our Media corpus: e.g. the 5-gram *by the national literacy trust* (the software by default works in case insensitive mode, that is ignores capital letters) occurs 35 times in 34 texts. Longer N-grams are a useful way of identifying set phrases, formulaic language but also longer titles and names.

The picture on the left shows 5-grams and their frequencies. There is another way of generating N-grams. The picture below shows 3-grams with one "open slot" indicated by the + sign.

	Type	Rank	Freq	Range
1	by the national literacy trust	1	35	34
2	of the national literacy trust	2	29	26
3	from the national literacy trust	3	28	27
4	diary of a wimpy kid	4	23	13
5	at the end of the	5	21	20
6	research from the national literacy	6	20	20
7	centre for literacy in primary	7	19	19
8	for literacy in primary education	7	19	19
9	the national literacy trust and	7	19	19
10	times more likely to read	7	19	18

	Type	Rank	Freq	Range	S1_TT	S1_Ent
40	there + a	38	228	138	0.057	0.507
41	to + children	41	226	158	0.35	0.857
42	in + the	42	223	151	0.57	0.948
43	their + and	43	219	160	0.598	0.951
44	in + to	44	206	143	0.34	0.795
45	a + for	45	201	139	0.642	0.95
46	the + literacy	45	201	106	0.085	0.2
47	more + to	47	200	138	0.265	0.543
48	national + trust	47	200	101	0.005	0.0
49	of + the	49	198	134	0.727	0.969
50	to + with	50	196	136	0.388	0.856
51	is + the	51	195	129	0.41	0.843
52	in + of	52	187	112	0.278	0.721
53	to + it	52	187	118	0.594	0.92
54	a + that	54	185	106	0.643	0.953
55	they + to	54	185	125	0.335	0.79
56	and + a	56	183	133	0.716	0.954
57	with + and	57	180	129	0.55	0.918
58	is + a	58	177	114	0.424	0.842

Search Query Words Case Regex N-Gram Size 3 Open Slots 1

The open slot (+) indicates that this position can have variable content. Line 46 (highlighted in picture) tells us that there is a variation in the trigram.

The next picture shows what the variation looks like. The 3-gram *the+literacy* reveals that while most of the occurrence account for *the national literacy*, there are also other possibilities such as *the physical literacy* and *the critical literacy*.

Left Context	Hit	Right Context
y chosen by the organisers – publishers Puffin and	the National Literacy	Trust – for the pilot programme. Called 'Puffin
nge of authors and texts from CLPE, NewsWise,	The National Literacy	Trust, BookTrust, The British Library and poet
naissance report was compiled in conjunction with	the National Literacy	Trust.
epresentatives from British Land, Drake Circus and	the National Literacy	Trust. A literary-themed prize will be
Opportunity Area is pleased to be able to support	the National Literacy	Trusts Inspiring Parents events. "We aim to
reads.org.uk. Allison Potter is the hub manager of	the National Literacy	Trust's Middlesbrough Reads. She has worked
Activity of Children and Young People. As chair of	the Physical Literacy	Action Research Group in Wales and strategic
ion Research Group in Wales and strategic lead for	the physical literacy	project in the region, she managed the
cheme to help increase their fitness and wellbeing.	The physical literacy	work undertaken by Jersey Sport at the
ociation and Google, we will help children develop	the critical literacy	skills they need to survive and thrive
allenge by helping primary school children develop	the critical literacy	skills they need to survive and thrive
their views on ways we can help with reading in	the community. "Literacy	Week has been amazing for the students.
ive was to implement the "national strategies" with	the daily literacy	hour and maths lesson: a highly prescriptive, "
ies 26 years shorter than children from places with	the fewest literacy	problems. We now know that our efforts
it. So what does McGeown's research contribute to	the gender literacy	gap? "In my recent studies I've
he median hourly waqe of workers in England with	the highest literacy	levels is 94 per cent higher than for

	Type	Rank	Freq	Range
1	reading for pleasure	1	64	42
2	reading and writing	2	48	31
3	reading writing and	3	23	19
4	reading age of	4	21	15
5	reading at home	5	12	9
6	reading is a	5	12	12
7	reading at the	7	10	9
8	reading age but	8	9	7
9	reading and literacy	8	9	9
10	reading habits of	8	9	9
11	reading in the	8	9	9
12	reading should be	8	9	9
13	reading for enjoyment	13	8	8
14	reading heroes competition	13	8	4
15	reading is the	13	8	8
16	reading and maths	16	7	5
17	reading comprehension strategies	16	7	5
18	reading to their	16	7	7
19	reading ability of	19	6	6

If we are interested in a particular word, we can also generate N-grams based on that term as in the picture on the left. Here we used *reading* as the word around which we wanted to see trigrams.

Tools

There are several ways of “doing” corpus linguistics. There are several specialised corpus linguistic online interfaces and several downloadable apps. Increasingly, those that have some coding knowledge use R and Python (this option is not going to be discussed further). Many researchers use combination of these as **every analytical tool has its affordances and limitations.**

One of the advantages of the online interfaces is that they offer a number of preloaded (often otherwise inaccessible) corpora to explore. Some of these corpora are very large (billions of words). SketchEngine, for example, not only offers state of the art interface but also a considerable number of linguistic resources (including other languages rather than only English). In addition to that it includes corpus building tools and an option to upload your own data to explore. SketchEngine is free to explore for one month (this information is accurate at the time of writing, March 2024).

Another excellent online interface **KonText** is provided by The Institute of the Czech National Corpus. The tool is free to use but you need to register. The interface has English version and it contains several resources in English and other languages, including the parallel corpus *InterCorp* for use in cross-linguistic research.

The best known downloadable apps, each with numerous functionalities, are **AntConc**, **#LancsBox** and **WordSmith**.

AntConc is developed by Prof. Laurence Anthony and is available for download here: <https://www.laurenceanthony.net/software/antconc/>. It is regularly updated and Anthony's website also includes links to short tutorials. A friendly tutorial can be also found at Heather Froehlich's webpage, this tutorial is specifically aimed at wider interdisciplinary audiences.

#LancsBox (<http://corpora.lancs.ac.uk/lancsbox/>), developed at Lancaster University, UK by Prof. Vaclav Brezina and colleagues, is a very powerful tool that has been recently considerably updated. Both AntConc and #LancsBox are **free to use**.

WordSmith (<https://www.lexically.net/wordsmith/>) has been one of the first tools, it's developed by Mike Scott and is now in its 9th version. Single User Licence costs 60 GBP. Each of the tools offers somewhat different interfaces and functionalities and it is largely dependent on individual preferences which tool you find the most suitable.

Analysis of Research Mobilities

In this section, we will show some snippets from our analysis of the Media and Twitter corpora as examples to illustrate the above described techniques and concepts.

Keywords and collocations in the Media corpus

We have generated a set of KWs from the Media corpus (as you have seen on page 53). As explained above, the set of KWs will depend on the statistical measure that is used and on the reference corpus, both need to be considered. The table shows KWs generated in #LancsBox with a much larger general newspaper corpus as a reference corpus. This comparison is based on the assumption that the KWs will show us specific words occurring in our Media corpus rather than words that are typical for newspaper corpus register generally.

One of the top ranking KWs is the word *reading*. This finding was consistent with findings in the other parts of the project, where 'reading' was one of the key themes of the primary literacy discourse. In order to see what subthemes emerge around the concept of 'reading', we generated a list of collocations (log likelihood statistics, span +/- 5). We decided to exclude function words (e.g., forms of the verb *be* and *have*, pronouns, determiners, numbers, frequent prepositions and conjunction and short adverbs such as *just* and *only*). We looked at 200 strongest collocates which we further divided into themes; selection of the themes that emerged are shown at the next page.



Ability	Age	Children	Comparison	Difficulty	Love & motivation
expected habits ability progress acquisition assessment standard levels tests level average developed abilities reached attainment	age ages year early years olds	children pupils child kids students boys	compared times better improved less higher development different increased greater	wars challenge help difficulties improve benefits challenging support problems improving	love pleasure enjoy enjoyment enjoyed fun enjoying motivation inspire encourage hero engagement heroes confidence encouraging passion

One of the two key themes that emerged from the thematically organised collocations of *reading* is ‘**ability**’ to read which is connected to children’s ‘**age**’. So we would find words like *acquisition, standard, levels, average, attainment* and *ages, year, early* (see the two first columns in the table above). See below how some of these words, here we illustrate *expected, ability* and *difficulties*, occur in the context of *reading* in the media discourse.

n to leave primary school in England having met the	expected	standard in reading,	writing, and maths by 2030. "Go
age 2 tests, 66 per cent of 11-year-olds reached the	expected	standard in reading,	and 70 per cent in maths. But
key stage 2 Sats last year, the gender gap at the	expected	standard in reading,	writing and maths was eight per
ed or disability (SEND), just 36 per cent reached the	expected	standard in reading,	and 37 per cent in GPS.So,
cent of children in Year 6 at the school reached the	expected	standard in reading,	with 5 per cent reaching greater
n 2012, the percentage of Year 1 pupils meeting the	expected	standard in reading	has risen from 58% to 82%, with
the subject. In 2017, 74% boys in Essex reached the	expected	standard in reading	at the end of the Key
the subject. In 2017, 74% boys in Essex reached the	expected	standard in reading	at the end of the â "
n 2012, the percentage of Year 1 pupils meeting the	expected	standard in reading	has risen from 58% to 82%, with
out than others: namely, those who didn't reach the	expected	standard in the	reading and grammar, punctuation a
country, one in four children are not reading to the	expected	standard by the	time they leave primary school. "Witl

The screenshot above shows a selection of concordance lines of *expected* as one of the collocates of *reading*. In this selection, we can clearly see the pattern *expected standard in/for/of reading*. Note, the word *standard* is also a collocate.

One of the dominant themes in the Media corpus is the “deficit discourse”, the next set of concordance lines of *difficulties* shows how the word *difficulties* occurs in the context of *reading*. Note that the word *not* is also on the collocates list.

Left Context	Hit	Right Context
ult to see local authorities placing children with reading	difficulties	in expensive placements when there are children who ha
lard assessment for identifying reading-comprehension	difficulties;	in fact, even for researchers, it can be difficult
er authority was denying the existence of children with	difficulties	in reading, or saying that they don't believe
r specific profiles of learning problems (such as reading	difficulties	in the absence of maths problems or low IQ),
ems. In research, children with reading-comprehension	difficulties	are identified using standardised tests. These tests typic
o be less likely to be diagnosed with dyslexia. "Reading	difficulties	are real. I've seen thousands of kids with
ts think that assessment and evaluation, where reading	difficulties	arise, may be areas where AI tools could really
e more likely to develop reading impairments, but such	difficulties	can't be uncovered until the child is three
e challenges that children with reading-comprehension	difficulties	face, there can be great difficulty in recognising these
are real. I've seen thousands of kids with reading	difficulties,"	he told me. "You know what? Very few of
etty good indication that they're having comprehension	difficulties."	Ideally, this diagnostic reading would take place one-to-
n who experienced persistent and unexplained reading	difficulties. 3.7%	of the children Yule surveyed on the Isle of
neurological basis, which affected bright children whose	difficulties	reading and writing could not be explained by the
ows us that many children with reading-comprehension	difficulties	stay below the radar or are inaccurately identified as

The *ability* to read is linked to ‘age’, see the following examples.

Left Context	Hit	Right Context
5-year-olds sitting National 4 and 5 exams have the reading	ability	of a 13-year-old or lower, a study into
year-old pupils sitting their National exams have the reading	ability	of a 13-year-old or lower, according to a
eaning that many pupils sitting their GCSEs have the reading	ability	of a 13-yearold or lower. This slump in reading
eaning that many pupils sitting their GCSEs have the reading	ability	of a 13-yearold or lower. This slump in reading
trail the rest of the UK in improving the reading	ability	of children, according to a study published today. The
trail the rest of the UK in improving the reading	ability	of children, according to a study published today. The
school from primary education with an assessed reading age	ability	of below their actual age in years. According to
he trend is apparent everywhere. Teenage children's reading	ability	is falling behind their chronological age. Last month a
sed to sophisticated stories on screen, so when their reading	ability	is holding them back books can make them feel
helping children to pick their next read – and that reading	ability	is only one factor that will influence this choice. "
all of these ideas may remain out of reach. "Inference	ability	is very important to reading comprehension and, in particular
book can be especially tricky when a student's reading	ability	is way out of line with the books dubbed
howed that there is a significant correlation between reading	ability	and GCSE results across all subjects. 'This was not
both boys and girls. "Consequently, the gap between reading	ability	and pupils' age grows every year of secondary school."

Another dominant theme we identified in the media discourse is related to ‘reading for pleasure’. The media discourse seems to heavily promote the idea that children either *love/enjoy* or should *love/enjoy* reading. See the following examples.

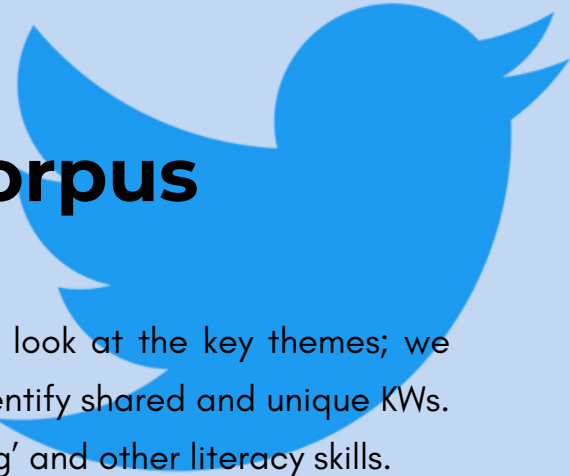
Left Context	Hit	Right Context
has really helped motivate the children to read more. "A	love	of reading is something that can last a lifetime.
and writing can be hugely beneficial. "Not only does a	love	of reading and writing enable children to flourish at
challenges, spending time with your child and encouraging a	love	of reading and writing is a good place to
and writing can be hugely beneficial. "Not only does a	love	of reading and writing enable children to flourish at
pleasure. Miss Miles added: "At Milton Park we encourage a	love	of reading and books from the minute they start
hopeless for the least well-off." Instead of encouraging a	love	of reading and writing, she says, schools have focused
selected in his book club which aims to foster a	love	of reading in disadvantaged children. Fifty thousand copi
literacy Toolkit containing 50 books and resources to inspire a	love	of reading in them. The toolkits were funded after
wonderland as part of a determined drive to encourage a	love	of reading in pupils. Every spare area has been
rough Reads and share our mission of encouraging a lifelong	love	of reading in children, get in touch by visiting
a chance at a prize), than to instil a genuine	love	of reading in a six-year-old who is
six, with the hope that all children will have a	love	of reading for life. "Teachers choose high quality books
for a Book initiative aims to help foster an early	love	of reading for children in a fun manner that

Looking back at the list of collocates (column 5 on page 58) you can notice the word *war*. We have initially placed the collocate *war* in the ‘difficulty’ column but upon closer examination, *war* occurs exclusively in *tes* and it is always in a phrase *reading wars*, a concept that will be, of course, known to specialists but is perhaps less clear for general audience, see below.

Left Context	Hit	Right Context
led us to draw two conclusions about why the reading	wars	have continued. First, phonics has been unfairly criticised, par
alphabetic writing system works. One reason why the reading	wars	have persisted is because of the view that children
unately, the vehement arguments and long-standing "reading	wars"	don't prove very useful for busy teachers looking
teachers with this endeavour, and to finally end the reading	wars.	Kate Nation is professor of experimental psychology at the
Perhaps the biggest casualty of the so-called "reading	wars",	or more specifically the fight over whether or not
to read. In this wide-ranging review, Ending the "reading	wars":	reading acquisition from novice to expert (bit.ly/ReadingWars)
standards (for more on this, see "Ceasefire in the reading	wars",	Tes, 15 June). The self-congratulation is warranted – we have
viewpoint, it is perplexing to witness the so-called "reading	wars" –	the controversies and debates about how reading should be
rs and confusions that have unnecessarily fuelled the reading	wars.	What follows is a summary of our paper, exclusively

Both KWs and collocations helped us to form an idea of our 400,000 word Media corpus without having to read every single article. This procedure would work equally well on much larger datasets. Both KWs and collocations are helpful pointers for further qualitative study.

Exploring the Twitter corpus



Similarly as with the Media corpus, we generated KWs to look at the key themes; we also compared the two KW lists – Media and Twitter – to identify shared and unique KWs. The shared KWs again emphasised the discourse on ‘reading’ and other literacy skills.

#Hashtags

Twitter, as a social media platform, has numerous specifics and conventions. One of them are **hashtags**. When working with hashtags, a useful step in the extraction is looking for variants in a hashtag’s spelling, for example, the hashtag indicating the *World Book Day* can be spelled as #worldbookday or #WorldBookDay. As these are clearly the same hashtag, it is reasonable to count them as one type. We looked at frequent hashtags to see which ones occur across the time span consistently and which are bound to specific points in time. The most consistently occurring hashtags are very general and unsurprising: #literacy, #education, #primary, #school, #reading. We also identified a considerable number of other hashtags relating to themes ‘reading’ and ‘book/s’, see below for a selection of the more frequent ones.

Hashtags 2019

#reading
#powerofreading
#readingforpleasure
#stokereads
#read
#dontleavereadingtochance
#read2change
#readingopensdoors
#bookmarkreadingcharity
#coderead
#gethastingsreading
#timetoread
#2019readingrevolution
#readingrocks
#gdnreading
#lovereading
#worldreadaloudday
#lovetoread
#readers
#readingisfundamental
#youngreadersprogramme
#1millionkidsreading
#youngreaders

Hashtags 2020

#reading
#primaryreadingpledge
#maskupandread
#readmcr
#coderead
#readingthefuture
#readgrowthrive
#readingforpleasure
#croydonyoungreaders
#summerreadingchallenge
#readingforsport
#readingrecovery
#readingchallenge
#timetoread
#readingcommunity

Hashtags 2021

#readingforpleasure
#reading
#road2reading
#readinginspires
#readaloud
#lovereading
#readingisfun
#scienceofreading
#everychildareader

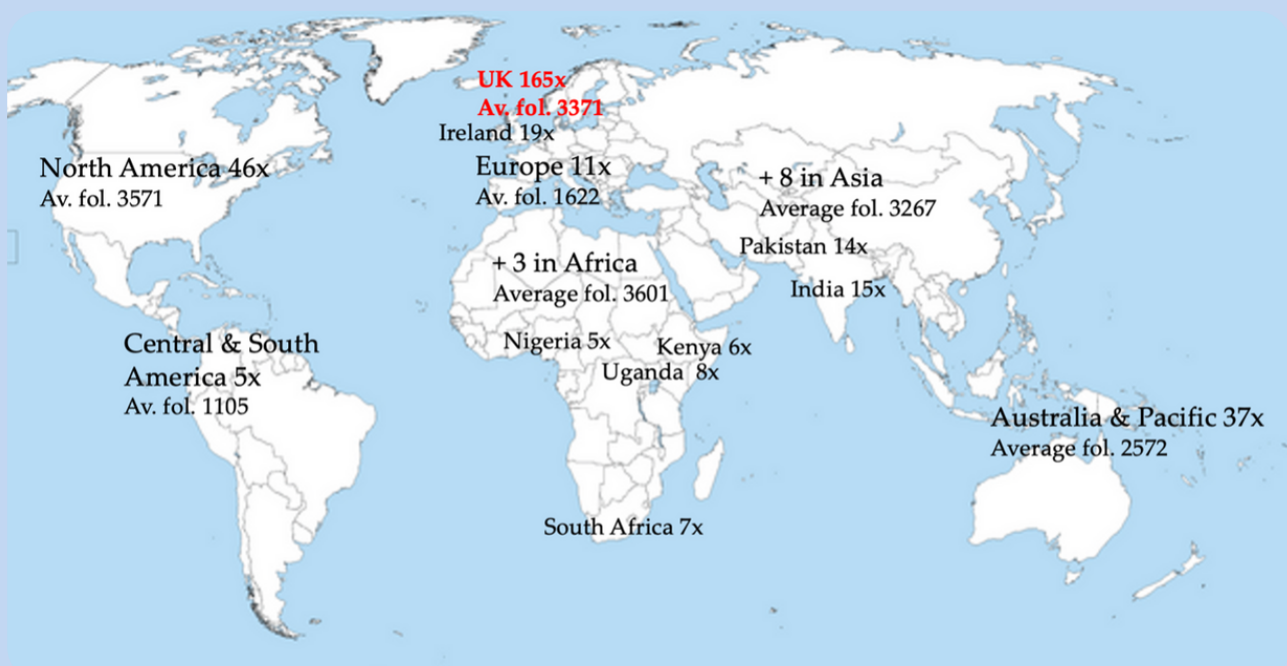
Hashtags 2022

#reading
#readingforpleasure
#storymojaread aloud
#readingcommunity
#thread
#worldreadaloudday2022
#readers
#readin
#lovetoread
#timetoread
#readingismagic
#readingpanel
#welovereading
#happyreading
#readerstodayleaderstomorrow

'university' accounts

One of the things we were interested in the Twitter data were individuals and organisations that were commenting on literacy in primary education. On Twitter, 'authors' (account holders) usually provide a short description of who they are. We were interested in 'academic' accounts (with the assumption they are more likely linked to 'research'). We decided to operationalise 'academic account' as an account that contains the word *university* in its description. We found 353 unique accounts that were distributed among individuals (250) and institutions (103). We then plotted these on a map, to see how they are distributed.

Please note, location is not a compulsory information on Twitter, some accounts include location information and some do not, and the way it is included is not standardised. Some manual analysis was therefore needed. The location was therefore determined in this order: 1) self declaration, 2) description, 3) if more than one location given, it is the current location rather than past (e.g. alumni). See the map below (produced in [Palladio](#)) for a summary version of the geographical distribution. It can be assumed that the skew towards accounts based in the UK is based on the fact that we collected the data in the UK.



After examining the location of academic accounts, we also explored in what sense they are “academic”. The word *university* would most frequently be referring to an affiliation of individual academics, a university or university department, library or similar and alumni, least frequently, it would be a student. Interestingly, there seemed to be different geographical trends, for example, among the UK based accounts, 32% would be indicating an affiliation of an individual academic, while alumni would be only 4%. There is an opposite trend for accounts based in Pakistan, only 7% of them belong to individual academics, while 43% indicate alumni.

Another interesting observation concerns the phrase *views my own*. While the academics are happy to declare their affiliation in their handle description, 55% also include some version of the phrase “views my own” in an effort to dissociate themselves from their institutions. These range from funny to matter-of-fact, see the examples:

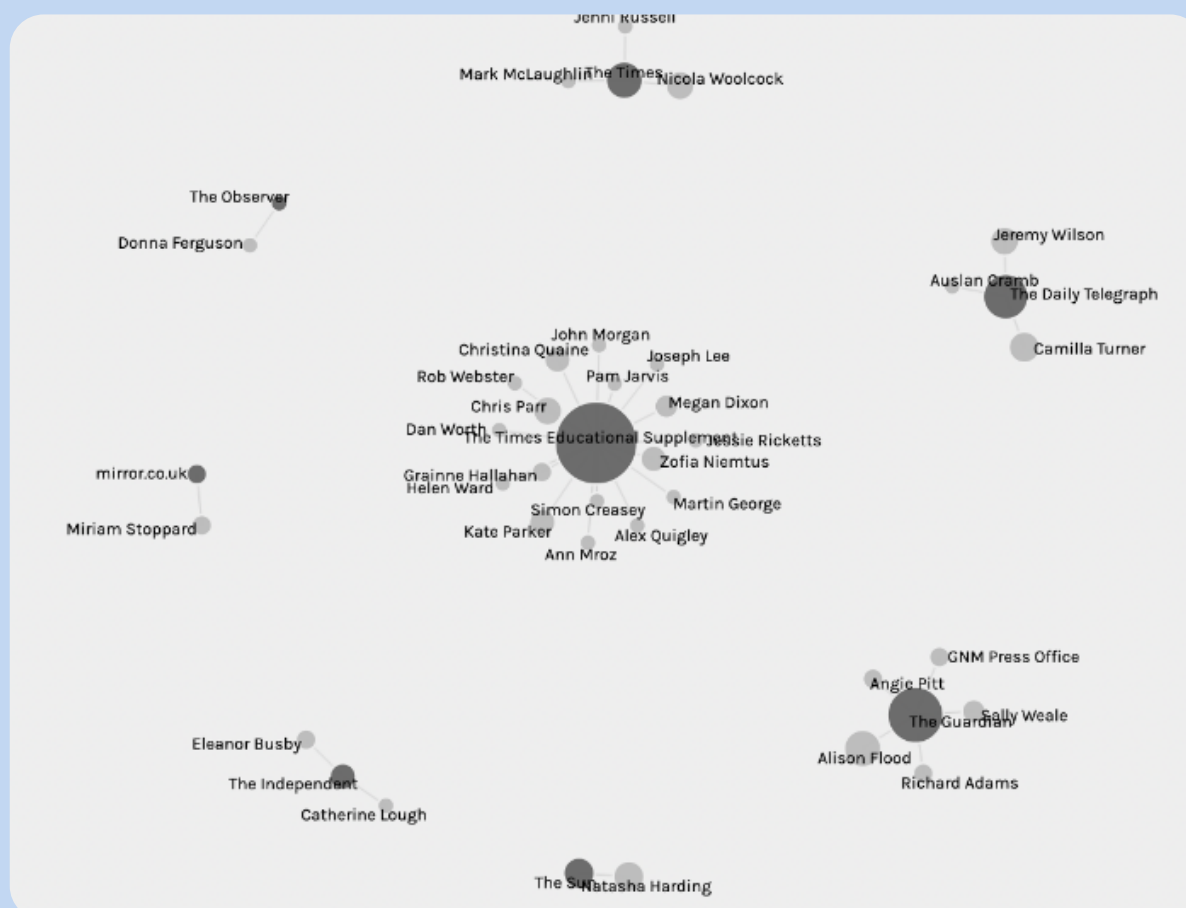
Opinions and Dad jokes are mine and not DoD

Views and opinions expressed are mine and not [...] University's.

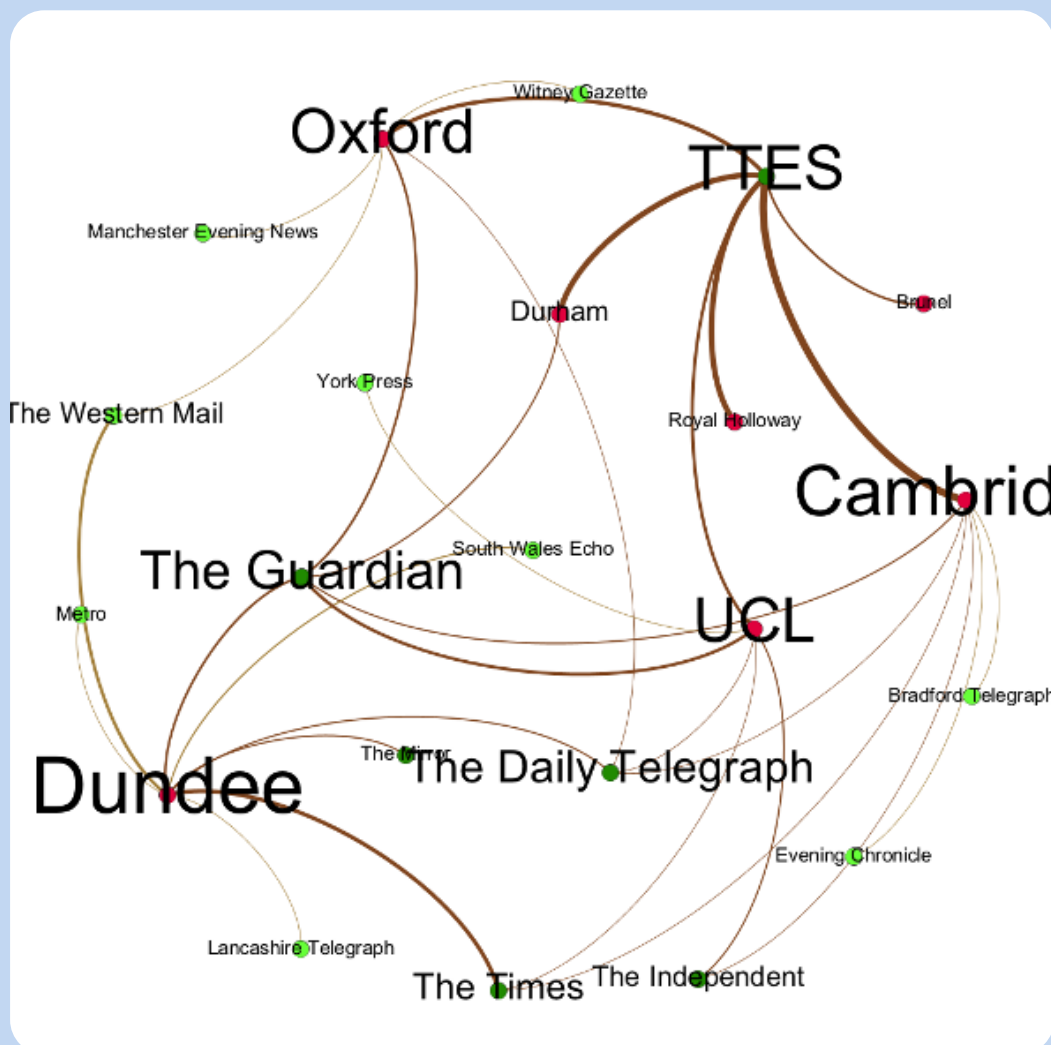
These are only some examples of analysis done on Twitter. Twitter as social media allows for numerous explorations that go beyond corpus linguistic methodology, network analysis would be a typical one. We were also interested in exploring the concept of influence in Twitter space. One way of looking at this, and thus considering the status and relative importance of a tweet in the discourse, is through retweets as explained above or other metrics, such as likes. Another would be to consider the accounts from which the tweet was sent in terms of the volume of their followers. One of the biggest accounts in our data was, for example, the account of the publisher Penguin House UK with well over 1.8 million followers. One of the surprises was the account of The Prince of Wales and The Duchess of Cornwall (username: @ClarenceHouse), now the King and Queen with about one million of followers. The account would, for example, inform its followers about Camilla reading to her grandchildren.

Visualisations

Data visualisations play increasingly more important role also in Corpus Linguistics. We used visualisations for different purposes. One of the first visualisations aimed to explore the media corpus structure in order to see the names of journalists (or other contributors) to newspapers. We plotted these into a tabular format and used the [Palladio tool](#) (free tool developed by Stanford University), the visualisation below shows journalists with two or more articles in the corpus linked to their newspaper. Some papers clearly have one or more journalists dedicated to education, while *tes* (in the middle of the picture) shows a great variety of names, many of which are not journalists but frequent commentators in the education discourse, such as Alex Quigley or Megan Dixon.



Another visualization we produced was while analysing the occurrence of the word *university* in the Media corpus. Here we again plotted these into a tabular format and used Gephi to visualize which newspapers write about which universities, see below.



Palladio and Gephi are tools developed within the larger context of Digital Humanities. However, as the importance of visualisations of textual data grows (see, for example the free online tool Voyant, which offers a number of visualisations and other analytical tools, however, cannot handle too large datasets), some of the corpus linguistic tools now also offer visualisation options. For example, #LancsBox produces collocation visualisation, see the figure on the next page, which shows collocations of the word *university* in our Media corpus.

Sociomaterial ethnography

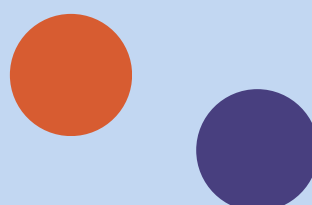
Terrie Lynn, David and Anna

Introduction

Our third methodological approach focused on tracing the “movements and translations” of specific “pieces of research”. ‘Research’ materialises in many different ways: in a familiar format of a journal article or published report with transparent research methods and findings, but, frequently, we encountered ‘research’ in myriad of other piecemeal formats: as a set of bullet points embedded within guidance, power point presentation, blog or social media post, a resource, a government guidance, a training session, or simply an idea or concept.

The broad theoretical anchoring in **sociomaterial methodologies** was inspirational and challenging at the same time. To help us navigate through these challenges we created a methodology influenced by **network ethnography, controversy mapping** and **more-than-human** approaches to ‘interviewing objects’.

We developed the methodology iteratively over several months through **nine cases**. Each of these cases had its own pace, began at different points in time and moved at different speeds with varying level of detail. The first one was, naturally, perhaps the most experimental, but we realised over the course of time with each one of them that a way of working developed for one case will not necessarily work for the next one without adjustments.



The cases were agreed based on the whole research team discussions, the teams' expertise and findings emerging from the work with the teachers. Cases were selected to maximise diversity across several dimensions. Initially, we wanted each of the cases to begin with a publicly available research text; however, the concept of 'research text' gradually expanded beyond published academic journal articles to also include research project websites, research reports, blogs and tweets, theoretical terms and concepts, and digital spaces in which students and teachers were enacting specific research.

The methodological approaches sketched in sections 'Qualitative approaches to teachers' encounters' and 'Corpus Linguistics' can easily adapt and incorporate various theoretical frameworks; we have, therefore, in this resource, intentionally largely left out our theoretical considerations. Leaving out some theoretical concepts is, however, largely impossible in this section and we will briefly outline the frameworks that we mostly drew on: Network Ethnography and Controversy Mapping. Both explicitly invite researchers to "follow the actors" (Latour 2005).

Network Ethnography

Network Ethnography (NE) entails identifying and analyzing the creation of networks and the connections that make up these networks. Network Ethnography also brings a strong focus on the work and labour that goes into how these networks come to be and what they do. Methodologically, it is very useful for studying mobilities, i.e., to see what is moving, does it stay the same, or does it change as it moves (see Ball 2016; Ball, Juneman & Santorini 2017).

In this project, we drew on a growing body of research employing NE, developed by Howard (2002) and later adapted by Ball (2016) to study **education policy mobilities**. In a recent 'state of the art' article, Ball (2023, p. 1) describes NE as a "responsive and adaptive assemblage of research tactics" necessary to study "the proliferation of policymaking sites and activities around the world and the increasing mobility and flow of education policy". Of interest is identifying and analyzing the creation and operation of global education policy networks and the connections that constitute them. Such networks are not just a set of connections between sites, but also represent a history of effort and various forms of materiality and performance (Ball 2016). Our focus in this project is on how various networks and actors of interest are inextricably intertwined, how is the network around a "piece of research" evolving, mutating, moving or splintering off into new networks.

In networks, actors assemble in certain ways and the configurations of these assemblages are animated by social and material relations and actions. These animations generate effects and enable certain entities, activities, actors, ideas, places, events, practices, and money to circulate in varying ways and degrees. Rowe (2022, p. 4) writes that NE is "exploratory, investigating, and scrutinising and seeks to reflect a critical lens; that is, it is epistemologically concerned with power and the struggle of power".

NE offers a way for researchers to attend to both the situated and the dispersed. Sometimes referred to as hybrid methodology, including **in-depth ethnographic work** as well as some form of **quantitative network-like analysis**. In our way of working with NE, we see this as much more than mixed methods (Robinson & Anderson 2022). Instead, we embraced the multiple and contradictory as we worked with network-like analysis alongside deeper ethnographic work, each with their own unique data insights and occasional and opportune intersections.

What is Network Ethnography practically?

NE entails a mix of methods, which may include: (1) deep and extensive internet searches, including social media trawls, blogs, podcasts, reports, speeches, white papers, organizational websites, printed curriculum guidance, online teacher training materials, practitioner online communities, and databases; (2) qualitative analysis of documents; (3) interviews and/or focus groups with key actors; (4) following key people or organizations on social media or blogs; (5) observation; and (6) the generation of visual maps. The degree of qualitative to quantitative work varies significantly across studies.

Controversy Mapping

Controversy Mapping enables researchers to take a closer look at some of the **relations and sticky points** within networks. Controversies are not always merely disagreements and arguments; they are generative events (Whatmore 2009), the effect of the uncertain and provisional nature of the production of scientific knowledge claims. Venturini and Munk (2022) describe controversies as **tensions which animate issues**.

Enabling observation of science in the making, CM (e.g., Munk & Ellern 2015, Venturini 2010, Venturini & Munk 2022) helps to foreground the (un)intended consequences of the ways science mobilizes, what it mobilizes, and the digital platforms which shape these mobilities (Venturini & Munk 2022). Such a focus aligns with our interest in research mobilities. Informed by Actor Network Theory (ANT) sensibilities, CM builds on the way it regards 'collective action' (Venturini & Munk 2022, p. 22).

Blackman (2019b) further opens up the notion and value of controversies. She questions what studying controversies can make visible, noting what she refers to as the “hauntological forces” that are not so easily mapped and likely to be obscured by digital methods based on semantic and content analysis (Blackman 2019b, p. 43). Attuning to what Blackman calls ‘**ghostly data**’ (p. 33) formed part of our work. Blackman (2019a) foregrounds the importance of more qualitative approaches in order to find and explore “submerged narratives, displaced actors and disqualifications that are often covered over, edited out, discarded or exist as minor agencies within legitimate science” (p. xxv).

The strong sociomaterial framing of CM and links to digital methods is well-suited to exploring the complexities of research mobilities that play out online or through interesting hybrid on/offline configurations. CM attempts to navigate a balance between data aggregation and more situated data analysis that focuses on the specific and particular interactions; describing itself as a “quali-quantitative method” (Venturini & Munk 2022, p. 146).

What is Controversy Mapping practically?

Controversy Mapping is very rich methodologically and very diverse in what specific approaches are employed. It depends on the controversy being studied and the resources available for the digital fieldwork. However, studying a controversy often moves through five stages asking questions: ***what, who, how, where*** and ***when*** (Venturini & Munk 2022):

What? *From knowledge claims to debates*

Who? *From debates to actors?*

How? *From actors to networks*

Where? *From networks to worldviews*

When? *From worldviews to change .*

Using similar approaches as Network Ethnography, Controversy Mapping works in innovative ways with mapping and building cartographies: a multitude of possible data visualizations enable multimodal analysis.

Both Network Ethnography and Controversy Mapping are influenced by sociomaterial perspectives such as **Actor Network Theory** and **Science and Technology Studies**. And so, attention is paid to things and objects as well as human actors. Here is where we turned to the notion of '**interviewing objects**' (Adams & Thompson 2016) to deepen both our attuning to and analysis of human-thingly interactions in our ethnographic fieldwork. As a strong more-than-human approach, object interviews focus on not only humans but also nonhuman things as relevant participants in social science research. In this project, this means focusing on how assemblages of such actors shape-shift research and move research together. Adams and Thompson (2016) provide several heuristics for the **more-than-human interviews**. In this resource, we highlight three of these heuristics: **following the actors** (building on Latour's (2005) work), gathering and writing descriptive **anecdotes**, and unravelling **translations** (Callon 1986).

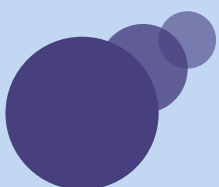
Translations

Callon's (1986) seminal work on translation describes 'translation' as a series of both displacements and transformations necessary to study power and the "structuring of power relationships" (p. 197). Translation enables researchers to examine how some research moves are imbued by powerful and influential actors along with specific manifestations or mutations of a piece of research. Translation examines how actors interface with others: willingly, under coercion, or unknowingly. It is through translations that entities meet up and interact with others, transform, become linked or de-coupled.

Callon (1986) originally outlined four moments of translation: **problemization**, **interessement**, **enrolment**, and **mobilisation**. Although presented in a rather linear fashion, progression through these four moments (which are not neatly bounded or entirely distinct) is not a certainty and assemblages and actors will jostle back and forth between moments as “margins of manoeuvre are negotiated and delimited” (p. 203). As Callon (1986) reminds, “translation is a process before it is a result” (p. 224).

We drew on translation as a conceptual heuristic to help us further untangle different research moves including the roles and actions of certain actors and changes to the research itself. Articulating various translations enabled us to glimpse how some understandings / positionings come to hold more power than others; which actors (and combination of actors) become influential and how; the nature of the connections and communications between actors; how stable any particular research assemblage is; and how research is enacted in particular ways and therefore contributes (or not) to research mobilities.

The first moment of translation, as originally described by Callon (1986), is **problemization**: an attempt to mark out or delineate an issue of interest which starts to link particular actors together. In our research project this may be how different actors gravitate to a particular manifestation of research. The devices of **interessement** are where the work really starts as actors become “interposed” (p. 208) when a “system of alliances” is constructed (p. 211). **Enrolment** is a further roping in of the actors in which particular roles are enacted and therefore enables a specific assemblage to take shape and begin to work. **Mobilisation** is a form of stabilization where the nitty gritty ongoing negotiations within the assemblage are “smoothed over” and the assemblage does what it needs to do and acts as “unit of force” (p. 216).



Following the actors...



Rather than a “connect the dots” approach to tracing research moves, we first attuned to appearances of the initial piece of research of interest. We then started to trace this research by following actors (human and otherwise) to explore what they were doing and what roles they might play in brokering research.

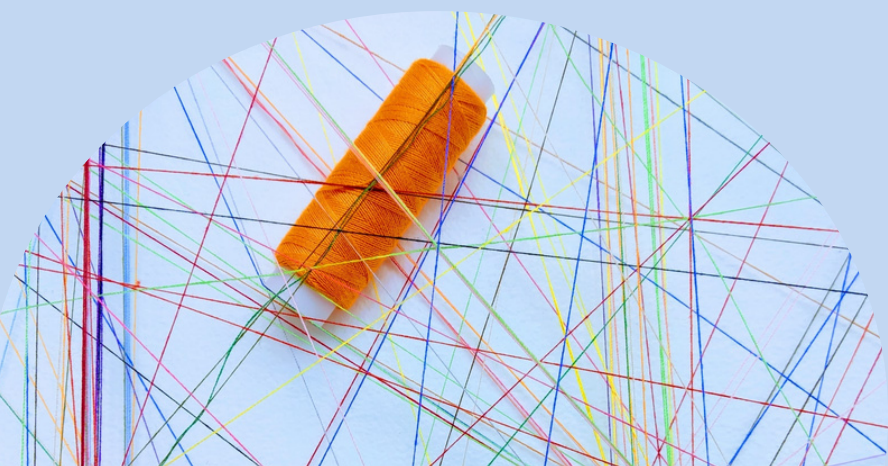
Following the actors means attuning to, tracing, searching for, and noting appearances of all kinds of actors, including ideas, places, events, practice, money, policies, activities, alliances, and organizations. It is equally important to know when to stop following some actors: to cut the network. That said, approaching the tracings with an openness to the unexpected, little anomalies, and possibilities within the seemingly everyday and mundane is what generates rich data.

After decision was reached on our cases, in order to follow the actors, it was essential to have **a starting point** regardless of how quickly each of these seemingly bounded “pieces of research” dis-assembled and re-assembled into a myriad of fragments and new configurations. However, as stated earlier, it is not a matter of connecting the dots but rather letting the appearance and movements of the research (or lack of visibility and traction) guide how the tracings unfold – which will necessarily be backwards, forwards, and sideways in time and space.

Selection of cases

Choosing nine cases was a challenging task not least given the wide diversity of literacy research we seemed to be touching upon in the project. We selected the cases to maximize diversity in terms of the following points that we felt were important to capture in one way or another:

- **absence or presence** within the data collected in the other parts of the project: some foci were chosen as they were present within the teachers' encounters data or news corpora and some because they were conspicuously absent;
- **literacy focus**: to reflect a range of kinds of literacy research (in terms of topics, methodology, stance). This was in line with our project focus on investigating the kinds of research that were mobilising and their currency, importance, and prominence in wider debates in England and elsewhere;
- **research brokers**: to examine the activity of a range of potential brokers that vary with respect to power and influence, credibility, depth and breadth of actions, and forms of labour they enact to mobilize research;
- **the presence of digital actors** and extent of digital mediation and traces: to take into account both prominent and less conspicuous digital spaces and technologies; the range of work undertaken by digital actors; the ethics and practicalities of accessing particular digital spaces and outputs; and the types of data they could generate;
- **the degree of planned versus more serendipitous actions by actors/brokers** that encouraged and facilitated specific movements of research;
- **temporality**: we selected research objects within origins in published research texts dating from the 1980s to 2023.



The nine cases

Phonics and the “reading wars”

Starting point:
published journal article
Wyse, D. & Bradbury, A. (2022).
Reading wars or reading
reconciliation? A critical
examination of robust research
evidence, curriculum policy and
teachers’ practices for teaching
phonics and reading. *Review of
Education* 10(1), e3314

Reading for Pleasure

Starting point:
the Open University website
(<https://ourfp.org/>)
& journal article:
Cremin, T. Mottram, M. Collins,
F., Powell, S. & Safford, K.
(2009). Teachers as readers:
Building communities of
readers. *Literacy* 43(1), 11-19.

Ofsted Review

Starting point:
Research
review series:
English
(Ofsted,
published May
2022)

Critical Connections

Starting point:
the project website:
<https://www.gold.ac.uk/clcl/multilingualism/criticalconnections/> &
<https://goldsmithsmdst.com/>

Reflecting Realities

Starting point:
*CLPE Reflecting Realities – Survey of Ethnic
Representation within UK Children’s Literature
reports* (by CLPE, 2018–2023)

Reading Fluency Toolkit

Starting point:
EEF Teaching and Learning Toolkit (online)

Funds of Knowledge

Starting point:
phrase and concept

Join the Dots

Starting point:
DOTS.team &
ARGLE.net websites

@OUPSecondary tweet on ‘The Reading and Vocabulary Project’

Starting point:
a tweet by @OUPSecondary posted in Feb 2022 from an account administered by the OUP

Starting points

As we gained a sense of some of the movements of specific research, we then selected various starting points for conducting different analyses. These are analytic starting points: our points of insertion into already existing swirls of research moves. The fluidity of our starting points is consistent with how various stakeholders encounter research. And because there are countless entry points, any research or research 'move' already contains moves and ideas from other times and spaces.

Each starting point was different ([see the previous page](#)). For the 'Phonics and the Reading Wars', it was a published research article, for 'Ofsted review', it was a published report. However, 'Reading for Pleasure', for example, had two starting points, a website and a published research article. Another case study had as its starting point a tweet that was identified during our [corpus linguistic analysis](#) of the [Twitter data](#). At times, the identification (and consensus) of what the starting point is, or should be, was far from clear, see the following David's reflection on 'Funds of Knowledge':

Our starting point for funds of knowledge is of research that is surprising by its absence. In this way, it is a case study marked by contradictions: both highly specific to a place (Tucson, Arizona) but with a significant international presence; both seemingly unknown in policy and practice contexts in England, and yet with evidence that it still circulates here in partial, submerged ways.

Unsure of where to start, I go to Google and search for "funds of knowledge" ... The search turns up 1.4 million results, straddling research, practitioner-facing blogs, and third space organisations. The first link is to an academic article by "LC Moll." Clicking the link opens a scan of an academic article in PDF form.

I open a separate tab, go to Google Scholar, and type “funds of knowledge” into the search box. The first result is “funds of knowledge for teaching,” by Luis Moll, Cathy Amanti, Deborah Neff, and Norma González. It has over 11,000 citations. I am familiar with Moll’s name. I click the PDF button: it loads the same PDF as the first link in the Google search, a 1992 article in *Theory Into Practice*. I read it. It’s only much later that I realise Google scholar is enacting some behind-the-scenes trickery: the Google scholar entry is for a reprint of this article in a 2005 edited collection, edited by three of the original article’s authors (Gonzales, Moll, and Amanthi), and which has itself amassed over 7000 citations. Yet, on Google Scholar, the entry for the 1992 reprint has somehow merged with its 2005 reprint. Its 11,000 citations seem to be spread across both the original publication and its reprint, and any attempt to separate them results in all manner of digital hijinks. (The irony that the second result is “an investigation of coherence” is not lost on me.)

Terrie Lynn’s experience was, however, different:

I begin by doing a simple Google search in my Firefox browser on ‘funds of knowledge’. In .28 seconds Google offers up 672,000,000 results. Where to start? Aware of Google’s page ranking algorithm I nevertheless decide to let it point me to what it considers the most visible iterations of Funds of Knowledge. The first result is a link to a pdf version of a poorly scanned version of the original Moll et al. (1992) research article posted by the University of California. OK – this is clearly a piece of research. Next in the list is a smartly packaged Funds of Knowledge Toolkit created by the Washington Office of Superintendent of Public Administration.

Clearly aimed at teachers, the toolkit includes sections on how to apply funds of knowledge to teaching and learning, how to identify students' funds of knowledge and how to integrate these funds into classroom practices (this section providing links to 'multiple examples of how fantastic teachers from across Washington described how they apply their students' funds of knowledge into their classroom lessons'). It ends with a template for recording students' funds of knowledge, broken down into more specific funds such as economic, geography, politics, sports, health, and cooking (and several more).

There are five references to academic research at the end. I return to my Google search and am interested in the link to the Funds of Knowledge Alliance, a recently-formed international organization (FundsOfKnowledge.org). The rotating banner at the top of the website catches my attention. One of the posts announces a newly published (2024) edited collection (Esteban-Guitart, ed.) by Routledge which includes chapters by members of this alliance. I note the currency of the publication and the large global scope with chapters focusing on both funds of knowledge and identity from Australia, China, the Netherlands, the Māori in New Zealand, Catalonia, Chile, Mexico, Vietnam, the US, South Africa and the UK.

It is on this website that I learn that Funds of Knowledge also goes by the acronym FoK. Curious I type 'FoK' into a new Google search and find that it means fill or kill (for equity orders), free of knots (Merriam Webster dictionary), and Forks over Knives (plant-based nutrition website). I smile and think how interesting the work of a network ethnographer is when one starts to follow the actors.

Many searches and refinements later, having traversed multiple digital spaces following up on promising leads; interesting detours; dead ends; reading hundreds of webpages, articles, and both texts and images; and discussions with research team members I have a better sense of its appearances of interest to us in England. For example, Funds of Knowledge appears in an article in *The Guardian*, published less than a month after the murder of George Floyd, Three ways for UK schools to improve their race relations now. But it also appears on X in discussions related to Argos catalogues (a UK retailer). It appears in spaces that span practice and scholarship, such as *Nursery World*, *Birth to 5 Matters*, and a blog from University of Nottingham's Primary Education Network. It makes at least 35 appearances on the *Writing for Pleasure* website. But it is also at the centre of a cluster of publications and research projects out of Tucson, Arizona. These include a 1992 academic journal article, written by Luis Moll, Norma Gonzalez, Cathy Amanti, and Deborah Neff (*Funds of Knowledge for Teaching*) that was reprinted in a 2005 edited book by Gonzalez, Moll, and Amanti (*Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*). In its preface, the book outlines its aspirations. It is 'for educators who are willing to venture beyond the walls of the classroom ... who are willing to learn from their students and their communities ... nurturing of students' strengths and resources as part of a 'call for greater teacher autonomy' (Gonzalez et al., 2005, p. ix). In this way, Funds of Knowledge is diverse in its permutations and subtle in its proliferations.

The two anecdotes highlight Blackman's (2019) notion of ghostly data, i.e., links or traces of "something else (of what we might call future potentials)" that may open to submerged and displaced actors and agencies (p. 32-33). As starting points were teased out, various routes opened and some other seemed more as dead-ends. For example in 'Critical Connections', we tested three different starting points, all were utilised in the first scoping stage but it was not until the first interview was conducted that the final starting point, the project webpage, was confirmed.

Data collection

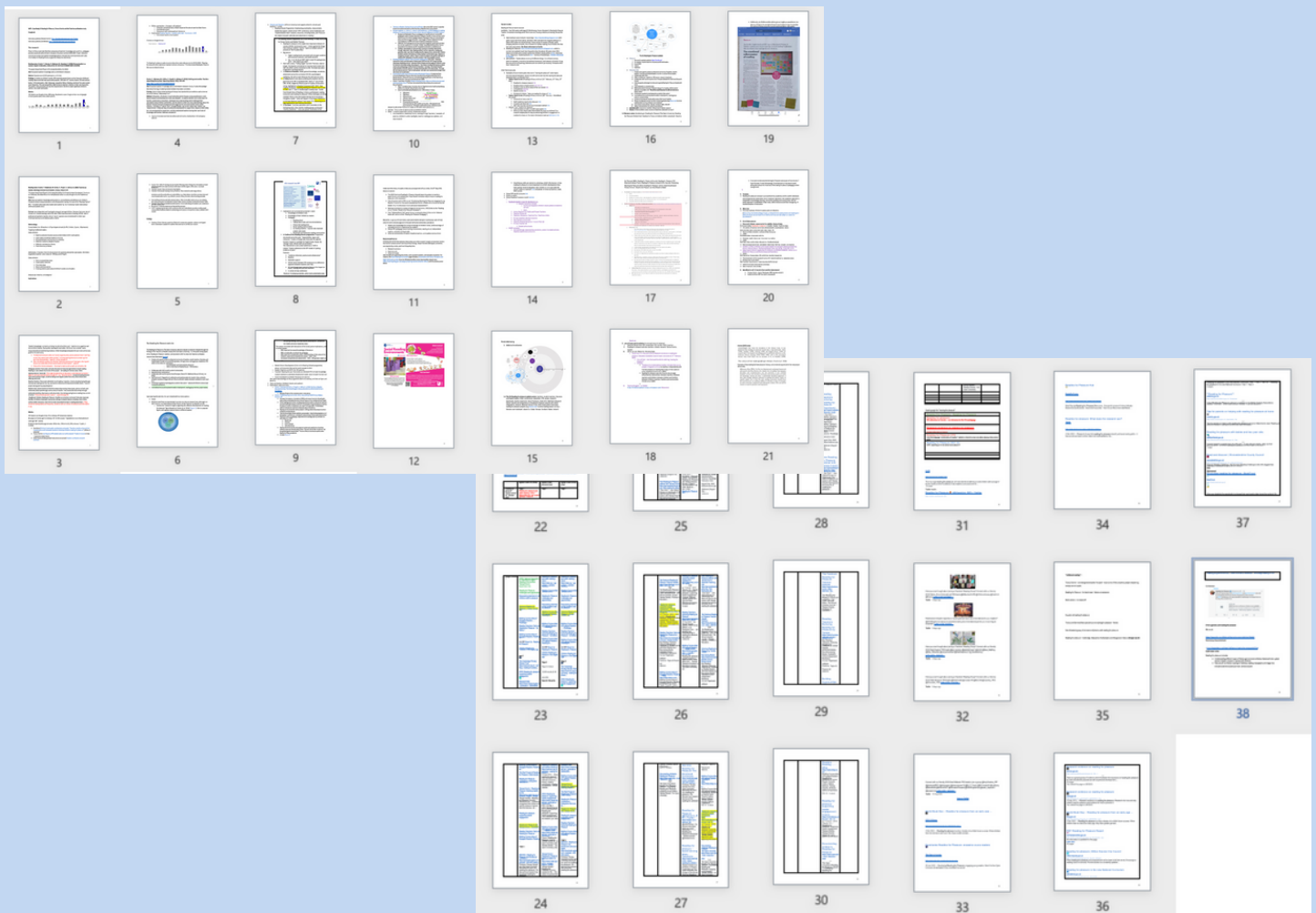
As the above shows, parameters for case study selection can be numerous and are difficult to generalise. Sociomaterial fieldwork for each tracing of a a case can take considerable time and resource. However, this can be managed by early and ongoing decisions on where to 'cut the network' and deciding on productive lines of movements to trace. That said, this is a form of **slow research** (Law & Singleton 2013).

As is perhaps apparent from the introduction, data collection was not straightforward and each of the cases required somewhat different data with multiple data formats. Data collection phase and analysis unfold together. The majority of the data was collected and captured digitally and online. In addition to 'tracings', we also collected data through interviews with people who we identified as relevant for our cases. We did a total of 18 interviews. Some of our cases included just one interview and for some we interviewed several people. The interviews were conducted on the Teams platform. They were recorded and then transcribed, if requested by the participants anonymised, and safely stored. Some of the interviews were relevant for several cases.

Relying on Google and other search engines, we were acutely aware of their **algorithmic politics**. The digital traces we encountered and interacted with are highly performative. We continually questioned **what we were seeing and why**. And **what we might not be seeing**. Because we worked online, we needed to keep detailed track record of the pages we visited and when, what we downloaded, when and where. Due to the fluid and constantly changing nature of internet, we would also frequently take screenshots or saved webpages, which is particularly relevant when working with digital data. Each of the researchers developed their own way of working to make this extremely time intensive data collection as efficient as possible.

Data collection and analysis unfolded in three different stages for each of the cases. The initial stage, **Initial Explorations**, consisted of time-limited period to formulate an “idea” of the case. After “organising” and discussing the first collected data set, another collection, **Stage 1 Analysis**, followed to expand on the previous one, this was followed by analytically oriented organisation of the data and further team discussions. The third collection, **Stage 2 Analysis**, period would include following up on specific points and conducting interviews. There is also a fourth phase which involved crating initial narratives based on the data. As we started to write these narratives, we returned to our data and in many instances, further pursued specific directions to gather more detailed data.

In some cases, for example, for the ‘Reading Fluency Toolkit’ by EEF, we have accessed 8 policy documents, 30 journal articles and 27 websites. In some cases, we created annotated “scrapbooks”, for example, the ‘Reading for Pleasure’ scrapbook contains 38 pages with 8455 words and numerous screenshots (see below for illustration), to hold data being generated, in other instances we created shared data folders.



Analytical framework

Initial Explorations

Aim

- Articulate a sense of what this research is/does and where/how it travels, in what forms and who some of the key actors/assemblages might be.
- Gauge depth/breadth and viability.
- Identify potential controversies (tensions).
- Identify the key brokers that could be a focus in this case study

Work on

- Collect basic information about the initial piece of research
- What is this piece research about, what is unique about it, what it is in response to, why this choice of publication or dissemination channel?
- What happened when this piece of research became public and how did it “appear”?
- Attune to the digital and thingly actors and their role in the mobility work of this research: what digital actors seem to be important?
- What translations did this research go through? What travelled intact? What became more diffuse? How was it reduced? Re-constituted?

Stage 1 Analysis

Aim

- Find and gather specific data: follow the actors in more in-depth strategic ways; explore possible data aggregations.
- Describe and do a relational analysis of actors including attuning to the labour of particular moves of the research.
- Begin to develop data anecdotes.
- Start to conduct interviews with human participants.

Work on

Actors and Assemblages

- Identify the longer list of actors of interest. Actors may include any of the following (and many others not in this list): consultants, web sites, practitioner communities, norm-setting models, conferences, guru performances, think tanks, blogs, research organizations, experts and lay experts, citizens and citizen-scientists, scientific campaigners and professional sceptics, digital actors, material actors in physical spaces, newspaper articles, hyperlinks extending across the internet, buzz phrases, policy aspirations, regulatory regimes, and specific discourses and public narratives.

Attuning to mobilities via the specific work and labour of particular moves

- Brokers: what actual work are they doing and how have they become involved in assemblings around this particular research?
- Examples of how the piece of research is being translated (Callon 1986).
- Identify and begin to develop more-than-human data anecdotes.

Analysis

Questions to stimulate analysis:

- Is new/different/the same kind debate being stimulated vis a vis the ways this research moves?
- Through the efforts, negotiations, and labour enacted in the various assemblages, what spaces for research are being enacted, traversed, opened up, closed down?
- Identify the potentially most fruitful actors and assemblages to further trace?
- Where do we “cut the network”?

Stage 2 Analysis: The Sociomaterial Analytic Framework

- What is moving and in what ways? What sort of knot-making (a temporary bringing together of energies) is going on?
- What does this research become as it moves? How does its meaning and impact shift?
- Identify tensions and/or controversies as well as any absences.
- Work through Callon’s (1986) four moments of translation for selected examples of translations.
- In what ways does this piece of research accrue status, influence and/or power? What relationships to other actors are making this possible?
- How is scientific expertise, legitimacy, and visibility of this piece of research being enacted? What mechanisms are in play to establish credibility and authority?
- How are digital (and other non-human actors) implicated in the displacement and transformation of this research? What is a digital technology explicitly or implicitly offering to help to frame thinking and intensify perception?
- How has this research (or its elements) become embedded in “material” things in day-to-day education literacy practices?
- Consider Blackman’s (2019) ghostly data: what becomes submerged or hidden by particular regimes of visibility?
- What patterns and disjunctures can be seen when we read across the cases?

Analysis

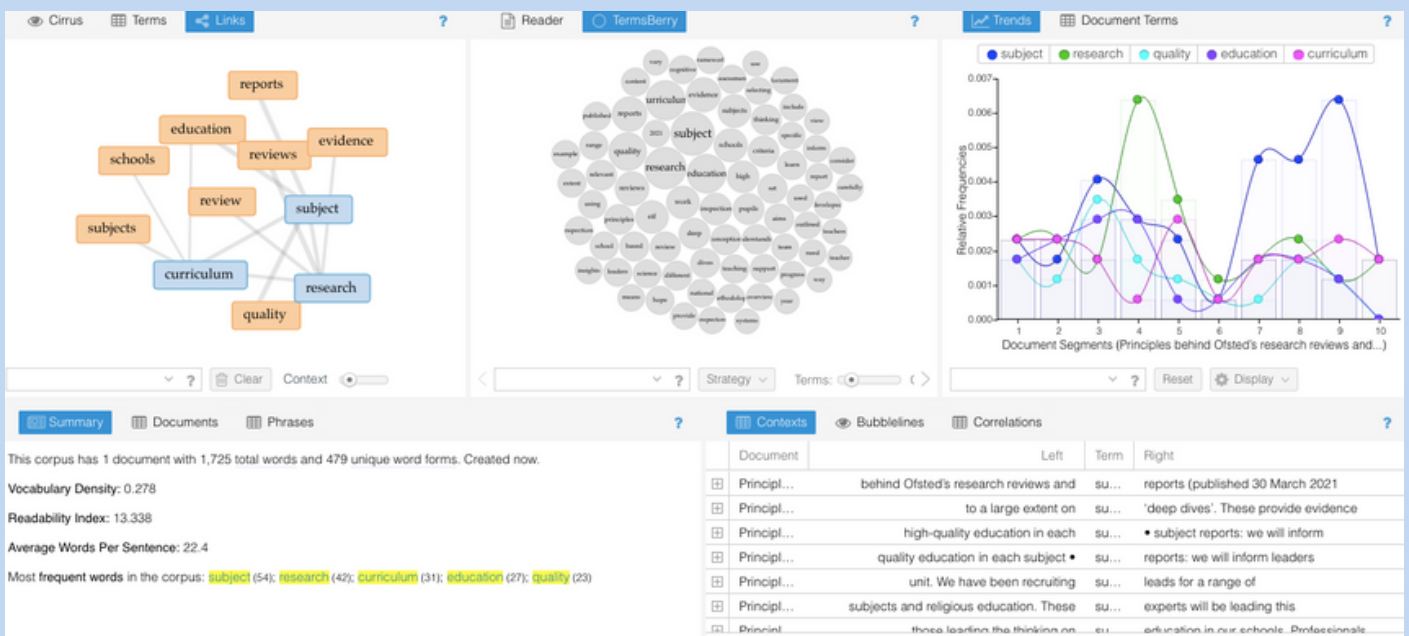
Tools for data aggregation

Our interest in qualitative digital tracing work notwithstanding, we were also interested and experimented, to some degree, with data aggregation methods, such as Google Trends (freely accessible), Media Cloud (Open source platform for media analysis, requires registration) and Issue Crawler (web network location and visualization software, requires registration). These powerful tools seemed, in many ways, to be “too powerful” in the context of some of our case studies where they did not have “sufficient” digital presence. We also felt we needed a certain level of detail which would be easily missed in the data aggregation; considering the project time constraints, we focused on more qualitative ways of exploring.



On the left, the screenshot from Google Trends shows a comparison of relative interest over time of ‘ofsted’, ‘phonics’ and ‘reading for pleasure’. While there seems to be consistent interest in ‘ofsted’ (with the blue trend line being more regular) and ‘phonics’, ‘reading for pleasure’ (the yellow line) barely registers.

For the text of the Ofsted report, we also experimented with the analysis in Voyant:



Anecdotes attend to the detail. They attune to the particular, the specific, what is happening in the moment, the concrete. Anecdotes strive to bring **invisibilities to the foreground**. They should convey a strong sense of what happens as things and people go about the everyday work of the world. Anecdotes serve as a heuristic to sharpen the sensibilities of a researcher to the things of everyday practice.

Although detailed, they are not merely a recital of endless details. Rather, anecdotes are compelling narratives that weave together both people and thingly storylines. The point of an anecdote should be evident. They detail what is happening, not why it happened. Anecdotes dig deeper than first glance descriptions of practice. As a researcher moves into analysis, anecdotes help to question and wonder, explore curiosities, and untangle what is happening in the unique interfaces and interactions between people and things.

CASE: Phonics and the “reading wars”

An article, published January 18, 2022, enters the ongoing debate about the ‘reading wars’ as made evident in its title. It is another empirical piece of research, based on a funded study, which argues for a more balanced approach to reading instruction in England.



Review of Education 

RESEARCH REPORT |  Free to Read

Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading

Dominic Wyse  Alice Bradbury

First published: 18 January 2022 | <https://doi.org/10.1002/rev3.3314> | Citations: 21

Funding information

Both authors benefit from funding from the Helen Hamlyn Trust as part of the funding for their research centre. No other funding is attributable to the work carried out for this paper.

Appearing as an article in a peer-reviewed journal, it is replete with the typical markers of academic work: publication date, a doi reference, journal name and publisher, its route to publication from receipt to revision to acceptance carefully documented with dates. There is a link to a YouTube video abstract featuring the two authors. Key words are carefully selected to tag the article in ways which will facilitate online searches, but also to signal the main aims of the article and align with specific debates: assessment, phonics, policy, and reading. The research article is already more than its 53 pages. Whatever this text is yet to become, it clearly does not work or travel alone: each of the actors just noted are implicated in how the article moves.

DETAILS | RELATIONS

BERA
Review of Education
Volume 10, Issue 1
Apr 2022

ARTICLE
Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading
[View article page](#)

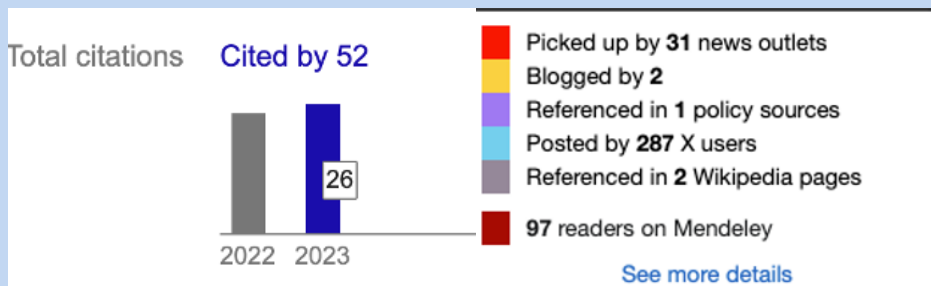
Dominic Wyse, Alice Bradbury

CITE | **Check for updates**

© 2022 British Educational Research Association.
<https://doi.org/10.1002/rev3.3314>

ISSN	2049-6613
eISSN	2049-6613
Online	18 January 2022
Accepted	17 November 2021
Revised	13 November 2021
Received	16 August 2021
Pages	n/a - n/a

Abstract




There are others. Metrics such as full text views (19, 546), an Altmetric score (460) (Oct 11/2023), and other digital actors, such as Google Scholar, help to count its reach (see Figure 9.3). The research article now rendered into a form that enables it to be datafied and tracked. Algorithms keep an eye on where it appears and influence how it might appear on screens in searches. Based on this data, there is clearly some movement as it is picked up, blogged or posted about, and referenced in policy sources or Wikipedia pages. And yet, despite all this information what exactly is being mobilized?

The screenshots from BERA website were taken on 3/10/2024, altmetrics info on BERA and Goolge scholar on 16/10/2023.

The case starts with the publication of Wyse & Bradbury (2022) article, which acquires an academic citation. The title and the construction of the **text of the article** provide further insights.

On page 4 the authors, Wyse and Bradbury say:

*The **effectiveness** of the teaching of reading **is** of **paramount importance** for education systems, and **effectiveness should** be determined through **rigorous** research. However, the extent to which relevant research evidence is reflected in curriculum policy and teachers' practice is affected by a range of influences. **For example**, political ideology can be an influence on the development of national curricula and teaching methods, sometimes contrary to the research evidence (e.g., as documented by James, 2012)*



The first sentence is in present indicative (the verb *is*) indicating the current state of things. There is strongly evaluative vocabulary (*effectiveness* 2x, *paramount*, *importance*, *rigorous*), leaving the reader in no doubt as to what is at stake. The next sentence, uses the modal *should* suggesting what the authors think is desirable. The following sentence provides an elaboration on why “effectiveness” may not be “determined through rigorous research”. One of the reasons, introduced casually by *for example* (suggesting there are (many) other reasons?), is “political ideology”. Thus, another line we seek to follow in this controversy is a power struggle behind different intellectual and ideological positions.

Various other devices which usually accompany publication of an academic journal article mark out an area of interest which starts to link actors together. It seems to be a typical academic article in a peer-reviewed journal. It is being read, written about, and cited. *But how does it move and where does it go?*

Adams and Thompson (2016) suggest a few object interview questions that can be posed to start the process of anecdoting. For example, “Describe how the object or thing appears or shows up.”, “What happens?”, “What happens next?”. These sorts of questions can be posed to human actors as well as digital things – key here is to get a sense of what is happening in the moment: the particular and specific moves of research and the actors and assemblages implicated in these movements and mutations to the research of interest.

Object Interviews: Unravelling translations



Another heuristic is to unravel translations. Translation enables researchers to examine how some research moves are imbued by powerful and influential actors along with specific manifestations or mutations of a piece of research. **Translation** is a powerful Actor Network Theory concept. Actor Network Theory asks: How has this collection of actors come to be assembled? Or disassembled? Or re-assembled differently?



Our data was filled with translations. For example, through the use of this heuristic we were able to study how research was translated into an event that could be read about; events that could be “attended”; and a controversy. Research was translated (e.g., displaced and transformed) as it materialized in film festivals and digital films created by school children and teachers (e.g. in ‘Critical Connections’). In some instances it was translated through a series of movements in which it became compressed into a sinuous concept with multiple agendas.

The research we examined in our cases experienced multiple translations and by using the notion of translation as a heuristic we were able to get a sense of the actors that draw together, around, and through particular research. How they form connections and alliances and in so doing, shape how research flows in particular ways and forms. Importantly, using translation analytically enables to see power and politics: what Callon (1986, p. 197) describes as the “structuring of power relationships”.

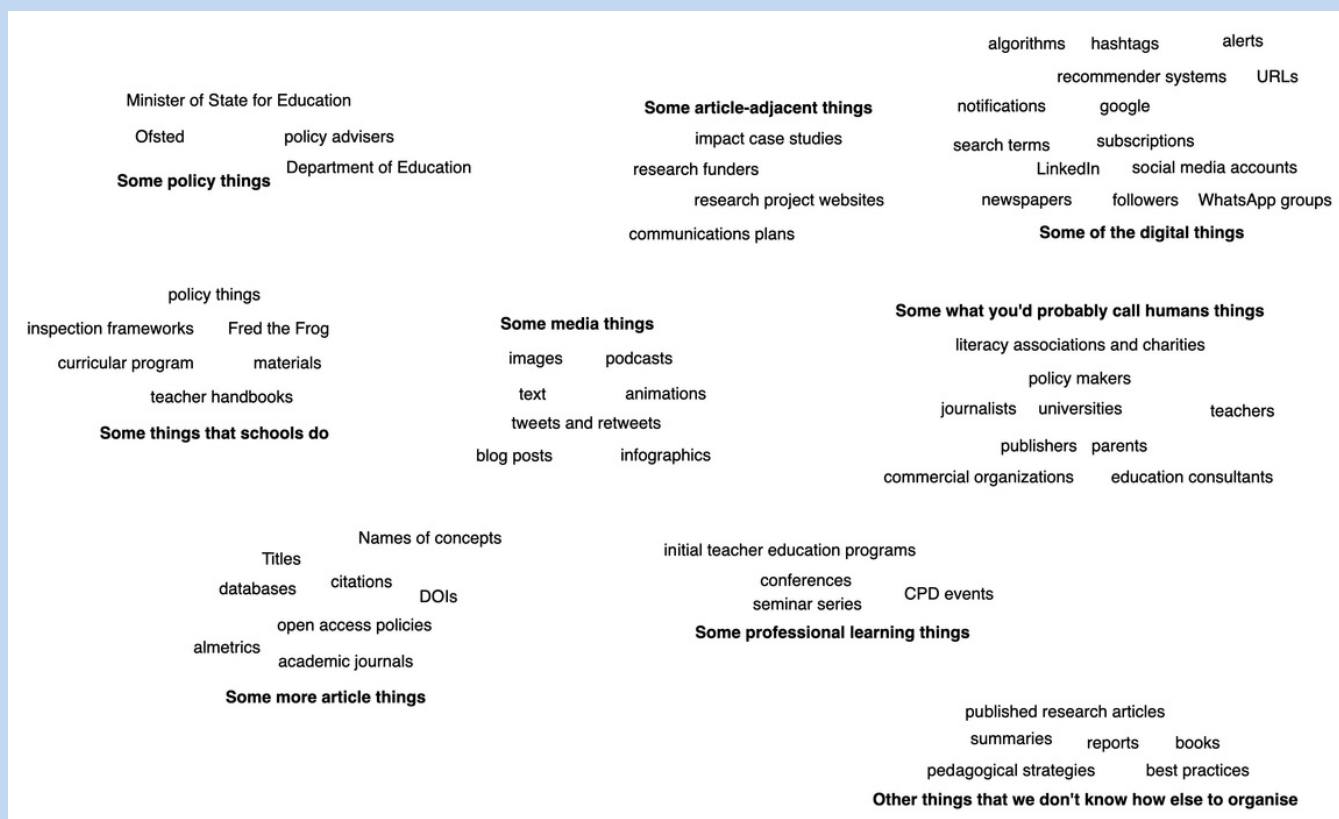
Mapping



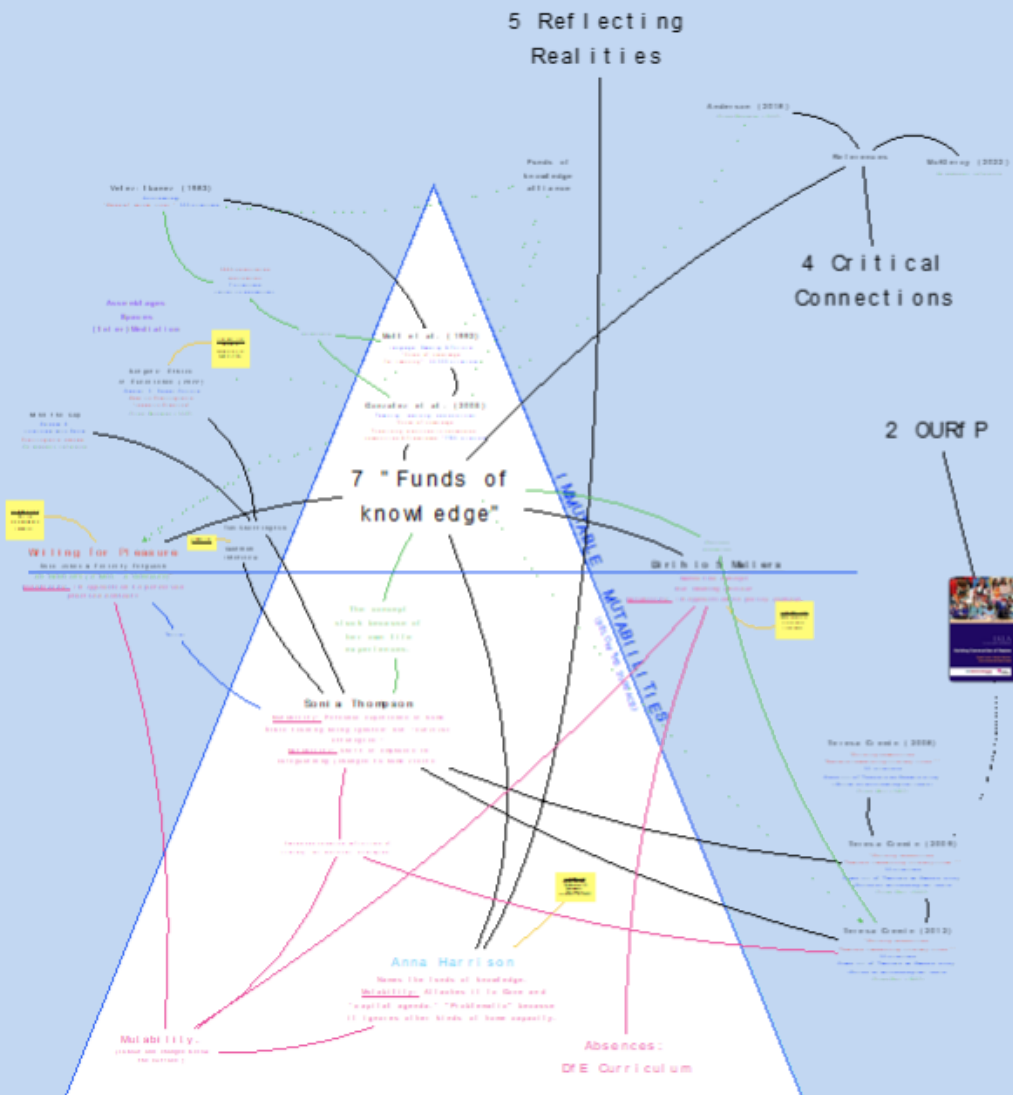
As mentioned above, the influences of Network Ethnography and Controversy Mapping took us deeply into data visualizations and mapping.

The aim, with the many ways we approached visualizations, was to try to bring not only networks into view, but the work and labour (including invisible labour) that animate these networks. We explored a range of visualizing approaches in an attempt to not “flatten” networks, which is often the case when someone (or a digital technology) tries to draw them out. Venturini and Munk (2022) advocate for the development of multiple “maps” and indeed, the **creation of cartographies and atlases** in order to convey the need for multiple, different kinds of visualizations as each map is necessarily partial. The generative potential of mapping “is not actually an attempt to represent the terrain, but an attempt to occupy the terrain and to engage with it” (Freitas & Walsham 2016, p. 4).

One of our attempts included visualisations of the types of actors we were following and mapping, we called it ‘Ephemera’:



CASE: Funds of Knowledge



The starting point for the 'Funds of Knowledge' case was the "phrase and concepts", see [David's](#) and [Terrie Lynn's](#) anecdotes above.

In the FoK case, the term and academic citation circulate as an 'immutable mobile'. Law and Singleton (2005), based on Latour, describe an immutable mobile as "something that moves around but also holds its shape" (p. 335).

For Law and Singleton (2005) the changes below the surface allow the "core of stability" to remain stable. This means that in practice, these flows below the surface (what ends up being blackboxed and translated out of view) alter the meaning of the concept vis a vis what is possible or permissible in the shifting curriculum and policy context. In other words, a constantly changing network of labour and relations enables the seeming stoicism of the stable relation between "funds of knowledge" and Moll et al. (1992; Moll 2005).

In the 'Critical Connections' case we became intrigued by the number and variety of actors moving in and out of this research project over 10 years. This made it difficult to describe all the ways in which this research continually re-assembles, sustaining itself as it both grows and decomposes. That said, the complexity of the aptly named 'critical connections' research focused our sociomaterial ethnographic fieldwork to attune to the many sites in which this research was being enacted, the connections that constitute the global networks which have formed, and the sorts of materiality implicated in the work and labour of research mobilities in this case. We provide an abridged version of our initial narrative about the array and work of actors that created or linked into different assemblages as a way to illustrate the visualization that we then generated.

The complex amalgam of funders and funding, partners, the range of mainstream and complementary schools in the UK and internationally, teachers, and students which have evolved over time. The CC 'research as project' or 'project as research' evokes a series of presences and absences that work to bring together different actors and galvanize various flows and circulations. As a research project, since 2012, CC has a history, supported by several funding sources (replete with the ongoing challenge and tensions of securing continued funding for the project).

*There is a steady publication output. Academic articles and blog posts continue to be generated by the project, each working to move the research by keeping its academic outputs visible and circulating. The open access **Handbook for Teachers** (Anderson, Macleroy & Chung 2014), is not only a handbook created for teachers, it also enacts a range of research-based experiences of working with teachers and students. In 2016, the **Multilingual Digital Storytelling** book was published with Routledge.*

Here is the Film Festival. Zoom and Eventbrite provide spaces and coordination of the online film festivals. But, perhaps most importantly for some actors, there is the activity in schools.

There are multiple entry points for actors gathering around, in, and through the CC project and the website. People, organizations, and an abundant ever-changing digital entourage seem to move in both coordinated and more organic ways. Social and material connections develop at events such conferences (including those hosted by the project in 2014, 2016, 2019, and 2022), annual film festivals, and planned partner networking and community engagement gatherings, including the Deptford Cinema.

Courses at Goldsmiths, CPD sessions, PhD studentships, and teacher and student digital story telling projects garner more connections between actors and shape specific research moves. Resources are rich and abundant, focused on clear support at all stages: pre-production-post. The legacy of schools and participants over an extended length of time who have engaged with this research become 'resources' for the project. As do the digital stories created by the students and faithfully archived by Vimeo, YouTube, and the website, amplifying reach through time and space and creating a visible history of effort and performativity of the research. Digital and physical spaces bring together digital and human actors; digital mediation facilitates the global reach of the project alongside local events which highlight the importance of specific locales.

The continued work in schools informed by, and informing this research along with the annual student entries in the international film festival and awards won by teachers give the successes of this research a presence. The relevance and popularity of (digital) storytelling, the interweaving of film making (based on the work of Joe Lambert in the USA), language learning and project-based learning, along with the creative process of drawing on one's own lived experience continue to be important mechanisms of articulation, persuasion and legitimation that give the research assemblage shape and focus.

None of these actors operate alone. And there are multiple overlapping networks. Callon (1986) refers to the negotiation of margins of manoeuvre, which, in this case, plays out across sprawling networks of resources, events, and pedagogical experiments.

International connections (for example schools in Taiwan, Italy, Japan, Germany, Cyprus, Malaysia, Australia, Egypt, India, Italy, Malaysia, Taiwan, Turkey, and Switzerland) and interactions between different stakeholder groups enter the mix. Notable are several high-profile partners: the British Film Institute, The British Museum and the Museum of London, and the National Resource Centre for Supplementary Education. The CC project, visible on some of these partner websites and the tangle of web links between elements of the research, is growing and becoming even more layered. An ethos of collaborative work and blurring the lines between research and practice frame the whole project.

Funding from several sources such as the Paul Hamlyn Foundation, the Qatar Foundation, and from Goldsmiths provide financial resources and expectations of certain kinds of 'deliverables' and outputs. The capabilities and capacity of the research team is important. The project team has grown, from Anderson and Macleroy to also include Dr Chung and Dr Rifesser; supported by project advisors, a support team and a social media manager. There is no one central mastermind (human or otherwise) directing actors and activities although there are several powerful and influential actors.

As a way of inspiring our thinking about connections and mobilities, we asked **ChatGPT 4.0** to provide a visualization of the narrative above. ChatGPT also offered a description. Acknowledging the idiosyncratic nature of any visualization and interpretation of it (especially when generated by an AI algorithm), the description evoked a lively discussion in our research team. For example, in its accompanying description, ChatGPT states that "at the center, a large, flourishing tree symbolizes the core research project". But is what we describe above such a symmetrical centered network? Various digital platforms mentioned above, "are depicted as floating platforms that link to the tree and to each other, highlighting their role in disseminating information and connecting participants across the globe" (ChatGPT).

Interesting that the digital is shown as somehow floating in space above people and other 'worldly' things. That said, perhaps this image does indeed "capture the dynamic interplay of people, ideas, and digital tools, all contributing to the growth and sustainability of the research project" (ChatGPT). We leave it to the reader to decide. With these sensibilities in mind, we found creative and academic value also in this thought experiment.



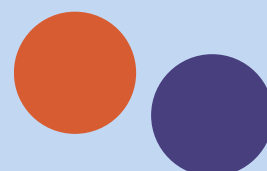
Conclusions

Ethics

Ethical approval for this project was given by Sheffield Hallam University Research Ethics Committee. The approval addressed participants' recruitment process, participants' consent procedures, confidentiality and data protection, which included data management plan. Ethical considerations were reviewed regularly at team meetings throughout the project and went beyond institutionalised procedures and requirements.

We were fully aware of our duty of care to project participants. Our project directly involved a wide range of participants in different strands of the study, both in direct and less direct ways, such as through the collected data. We worked directly with teachers (interviews, focus groups and lifelogging), we held stakeholder panels, and, as part of our sociomaterial ethnography inquiry we conducted interviews. We also encountered participants indirectly in the digital online space. The diversity of participants raised diverse potential ethical concerns and, at all times, we treated the participants' contributions with considerable care: this involved, for example, providing opportunities for participants to redact parts of interview transcripts and adhering to university protocols for data security and management.

In data transcriptions, our default position was to pseudonymise transcripts and extract details of places or institutions that would make individuals identifiable. However, with the ethnographic interviews conducted for our cases, we offered the interviewees the chance to be identified if they wanted to be associated with the work we were describing. In addition, in all these cases, we operated considerable caution in selecting extracts of material to share through publications or presentation.



Reflections

The combination of the three methodological vantage points we adopted in our project presented a fertile ground for reflections across the team. Research mobilities, as any research really, are tangled up in a multitude of personal, political, social, cultural and ideological landscapes. Inevitably, our own positionality as researchers, with our varying personal, social and cultural backgrounds has in many ways influenced the adaptation of methodologies, interpretation of the collected data and research outcomes. It is also likely that we were not always fully aware of these positionalities.

The sociomaterial ethnography work particularly highlighted the larger issues inherent in digital research. We were, at all times, aware of the underlying algorithms that influence our digital tracing. The underlying question being – how would have our findings been different have we searched in another country on someone else’s device at a different time using a different search engine? Similar concerns arose during the Twitter data collection. How representative the data actually really was? How did the Twitter algorithm decided to select the data for us? Soon after we collected the Twitter data, Twitter became X – and we posed a question, would we have still chosen the X data for exploration?

Despite the considerable amount of uncertainties, we feel that methodologically, we were able to find our unique angle into research mobilities and we sincerely hope this resource will be valuable to other researchers on their exploration journeys.



REFERENCES

- Adams, C. & Thompson, T. L. (2016). *Researching a Posthuman World: Interviews with Digital Objects*. Palgrave Macmillan.
- Ball, S. J. (2016). Following policy: networks, network ethnography and education policy mobilities. *Journal of Education Policy*, 31(5), 549–566.
- Ball, S. J. (2023). Education policy network analysis: The state of the art. *International Journal of Education Research*, 117, 1–2.
- Ball, S. J., Juneman, C. & Santorini, D. (2017). *Edu.net: Globalisation and Education Policy Mobility*. Routledge.
- Blackman, L. (2019a). *Haunted Data: Affect, Transmedia, Weird Science*. Bloomsbury.
- Blackman, L. (2019b). Haunted data, transmedial storytelling, affectivity: Attending to ‘controversies’ as matters of ghostly concern. *Ephemera: Theory and politics in organisation*, 19(1), 31–52.
- Braun, V. & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352.
- Braun, V. & Clarke, V. (2022). *Thematic Analysis: A practical guide*. SAGE.
- Brezina, V., Hawtin, A. & McEnery, T. (2021). The Written British National Corpus 2014 – design and comparability. *Text & Talk*, 41(5–6), 595–615.
- Burnett, C., Adams, G., Gillen, J., Thompson, T. L., Cermakova, A., Shannon, D. & Shetty, P. (2024). *Research Mobilities in Primary Literacy Education: How Teachers Encounter Research in An Age of Evidence-Based Teaching*. Routledge.
- Burnett, C., Gillen, J., Guest, I., Maxwell, B. & Thompson, T. L. (2022). How does research reach teachers? *Literacy*, 56(4), 386–399.

- Burnett, C. & Merchant, G. (2020). *Undoing the Digital: Sociomaterialism and Literacy Education*. Routledge.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, Action and Belief: A New Sociology of Knowledge?* (pp. 196-233). Routledge & Kegan Paul.
- Chadwick, A. (2017). *The Hybrid Media System: Politics and Power* (2nd ed.). Oxford University Press.
- Collins, L. & Hardie, A. (2022). Making use of transcription data from qualitative research within a corpus-linguistic paradigm: issues, experiences and recommendations. *Corpora*, 17(1), 123-135.
- Freitas, de E., & Walshaw, M. (2016) Introduction. In E. de Freitas & M. Walshaw (Eds), *Alternative Theoretical Frameworks for Mathematics Education Research: Theory Meets Data* (pp. 1-9). Springer.
- Frigo, A. (2016). As We Should Think? Lifelogging as a Re-emerging Method. In Kubitschko, S. & A. Kaun, A. (Eds), *Innovative Methods in Media and Communication Research* (1st ed., pp. 139-159). Springer International Publishing.
- Howard, P. (2002). Network ethnography and the hypermedia organization: new media, new organizations, new methods. *New Media & Society*, 4(4), 550-574.
- Joulain-Jay, A. (2021). *Tweet Collector V4.01* [Computer Software].
- Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-network Theory*. Oxford.
- Law, J. (2004). *After Method: Mess in Social Science Research*. Routledge.
- Law, J. & Singleton, V. (2015). Object Lessons. *Organization*, 12(3), 331-355.

- Law, J. & Singleton, V. (2013). ANT and Politics. Working In and On the World. *Qualitative Sociology*, 36, 485-502.
- Love, R., Dembry, C., Hardie, A., Brezina, V. & McEnery, T. (2017). The Spoken BNC2014. Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319-344.
- Mason, J. (2002). *Researching Your Own Practice: The Discipline of Noticing*. Routledge Falmer.
- McCormack, C. (2004). Storying stories: a narrative approach to in-depth interview conversations. *International Journal of Social Research Methodology*, 7(3), 219-236.
- Mol, A. (2003). *The Body Multiple. Ontology in Medical Practice*. Duke University Press.
- Mol, A. & Law, J. (2002). Complexities: An Introduction. In J. Law, A. Mol, B. Smith & E. Weintraub (Ed.), *Complexities: Social Studies of Knowledge Practices* (pp. 1-23). Duke University Press.
- Moll, L. (2005). Reflection and Possibilities. In N. Gonzalez, L. C. Moll & C. Amanti (Eds), *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*. Lawrence Erlbaum Associates.
- Moll, L., C., Amanti, C., Neff, D. & Gonzalez, N. (1992). Funds of knowledge for teaching: using a Qualitative approach to connect homes to classrooms. *Theory into Practice*, 31(2), 132-141.
- Munk, A. K. & Ellern, A. B. (2015). Mapping the new Nordic issue-scape: How to navigate a diffuse controversy with digital methods. In G. T. Johannesson, C. Ren, & R. van der Duim (Eds), *Tourism Encounters and Controversies: Ontological Politics of Tourism Development* (pp. 73-96). Ashgate.
- Robinson, S., & Anderson, C. W. (2020). Network Ethnography in Journalism Studies: A Mixed-Method Approach to Studying Media Ecologies. *Journalism Studies*, 21(7), 984-1001.

Rose, G. (2016). *Visual Methodologies: An Introduction to Researching with Visual Materials* (4th ed.). SAGE.

Rowe, E. (2022). The assemblage of inanimate objects in educational research: mapping venture philanthropy, policy networks and evidence brokers. *International Journal of Educational Research*, 114, 1-13.

Saldaña, J. (2021). *The Coding Manual for Qualitative Researchers*. SAGE.

Scott, M. (2009). In Search of a Bad Reference Corpus. In D. Archer (Ed.), *What's in a Word-list? Investigating Word Frequency and Keyword Extraction* (pp. 79-92). Oxford: Ashgate.

Scott, M. (2010). Problems in investigating keyness. In M. Bondi & M. Scott (eds.), *Keyness in Texts* (pp. 43-57). John Benjamins.

Scott, M. & Tribble, C. (2006). *Textual Patterns: Keyword and Corpus Analysis in Language Education*. John Benjamins.

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science*, 19(3), 258-273.

Venturini, T. & Munk, A. K. (2022). *Controversy Mapping. A Field Guide*. Polity.

Whatmore, S. J. (2009). Mapping knowledge controversies: Science, democracy and the redistribution of expertise. *Progress in Human Geography*, 33(5), 587-598.

Interactive index of key terms and methods

anecdotes

AntConc

artistic responses (visualisations)

coding

collocation

concordance

Controversy Mapping

corpus linguistics

data aggregation

distribution of words (in a corpus, 'plots')

Excel

focus groups and interviews

following the actors

ghostly data

hashtags

keywords

Lancsbox

lifelogging

mapping

media (newspaper) corpus

n-grams

Network ethnography

'noticing'

NVivo

object interviews

reference corpus

teachers' 'mentions'

stacking stories

storying

thematic analysis

themes (generating)

translation (Callon 1986)

Twitter corpus

word frequency

WordSmith

workshops

Acknowledgment



This work was supported by the
Economic and Social Research Council -
grant no. ES/W000571/1

How to cite:

Cermakova, A., Adams, G., Burnett, C., Gillen, J., Thompson, T.L., Shannon, D.B., Shetty, P., & Vackova, P. (2024). *Investigating Research Mobilities: A methods resource*.

<https://doi.org/10.7190/shu-reports-24008>

Sheffield Hallam University

Investigating research mobilities: A methods resource

ĀERMÁKÓVA, A., ADAMS, G., BURNETT, Catherine <<http://orcid.org/0000-0001-6087-244X>>, GILLEN, J., THOMPSON, T.L., SHANNON, D., SHETTY, P. and VACKOVA, P.

Available from the Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/33376/>

Copyright and re-use policy

Please visit <http://shura.shu.ac.uk/33376/> and <http://shura.shu.ac.uk/information.html> for further details about copyright and re-use permissions.