Department of Education

UNIVERSITY OF BATH

Ethical considerations & methodological issues in the creation of a corpus of young learners' disciplinary writing

Dr Reka Jablonkai & Professor Gail Forey
Department of Education
University of Bath

BAAL Corpus Linguistics SIG Symposium
Exploring the Use of Corpus Linguistics Research Methods

Wednesday 27th March 2024, online Zoom

Outline

Aims and rationale

Ethical issues in relevant fields

Ethical challenges, considerations and solutions

Methodological considerations: Corpus design and preparation

Conclusions

Aims and rationale

BAWESS (British Academic Written English of Secondary School) Corpus

a discipline-specific corpus of authentic student exam-practice written texts collected from UK and international schools

Little literature with systematic discussion of ethical issues and considerations in corpus building, corpus sharing (e.g. Leedham et al., 2021; Lillis et al., 2023)

Few corpora of school texts (e.g. Durrant & Benchley, 2018; Hamid & Crosthwaite, 2022)

Hard-to-access texts

Under-age writers

Under-researched genres

Ethical principles and guidelines

- 1 Inclusivity
- 2. Respect
- 3. Integrity
- 4. Social responsibilities in conducting and disseminating their research.
- 5. Aim to maximise benefit and minimise harm.

(Academy of Social Sciences, 2015)

Relationship and responsibilities – BAAL Good practice (2021), BERA Guidelines (2018)

macroethics – microethics Kubanyiova (2008), De Costa (2015, 2024)

procedural ethics – ethics in practice (Lillis et al., 2023)

Transparency, confidentiality, reflexivity, complexities, ethics training, inequities of authorship, interpreting test scores (De Costa et al., 2021)

Ethical issues in corpus research

Ethical issues arise from corpus construction, distribution and use

Relating to respondents, distributors, users of corpus data (McEnery and Hardie, 2012)

Public and private distinction of texts (McEnery & Brookes, 2022)

Sarangi (2019) in Lillis et al. (2023) WiSP corpus

Ethics of access

Ethics of participation

Ethics of interpretation/ representation

Ethics of dissemination/intervention

Ethical considerations in the case of the BAWESS corpus

Access to text writers

Ethical considerations around working with young persons

Metadata, jigsaw information

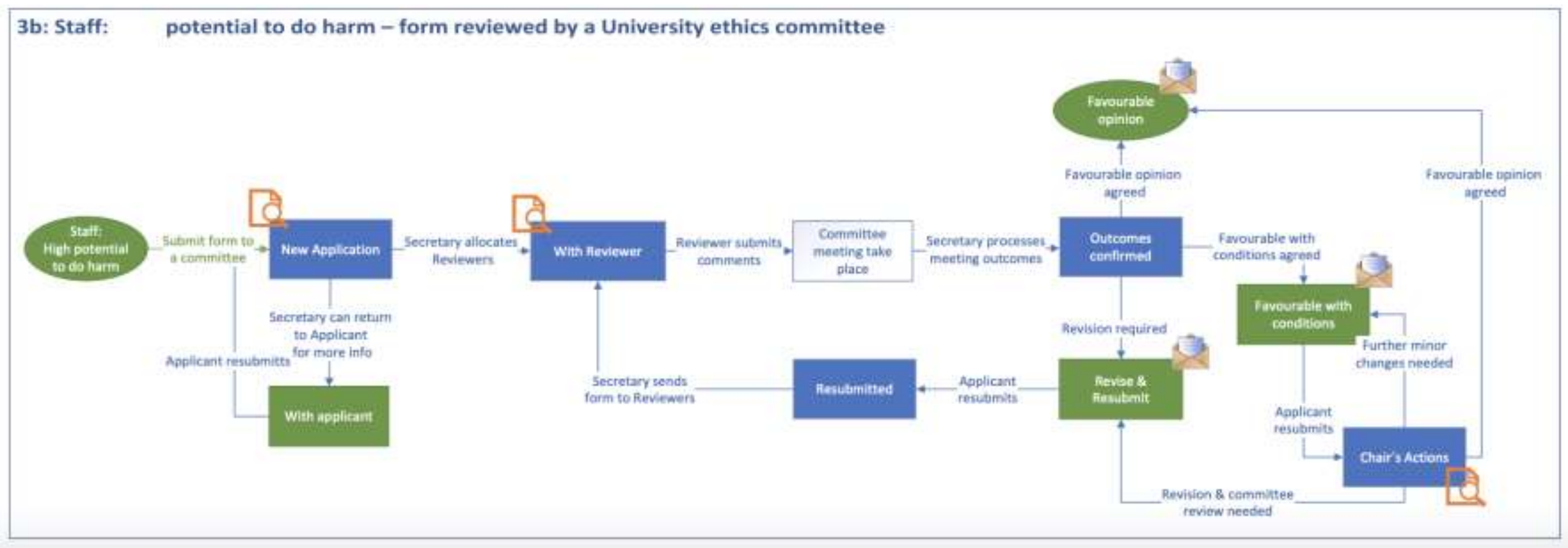
Use, collection, and preparation of texts

Access to corpus for research purposes

Access to corpus for instructional purposes

Institutional ethical approval

Ethical approval
PIS, Consent Forms, U16 Assent Forms & Parental Consent


3b: Staff: potential to do harm – form reviewed by a University ethics committee
Staff: High potential to do harm
Submit form to a committee
New Application
Secretary allocates Reviewers
With Reviewer
Reviewer submits comments
Committee meeting take place
Secretary processes meeting outcomes
Outcomes confirmed
Favourable opinion agreed
Favourable opinion
Favourable with conditions agreed
Favourable with conditions
Favourable opinion agreed
Further minor changes needed
Applicant resubmits
Chair's Actions
Revision & committee review needed
Revision required
Revise & Resubmit
Applicant resubmits
Resubmitted
Secretary sends form to Reviewers
Secretary can return to Applicant for more info
Applicant resubmitts
With applicant

Consent

Age of writers

Anonymity

Metadata

Digitalisations of data

Sharing & publishing

Corpus development: Key Considerations, Challenges & Solutions

Co-creation

Resources Time

Consent: Ethics

Knowledge, value & buy in

Co-Creation & Collaboration

Value – mutual benefits

Finding collaborators, building partnerships, invest time in building pool of collaborators

Negotiating the collaboration

Contacting and working with gate keepers

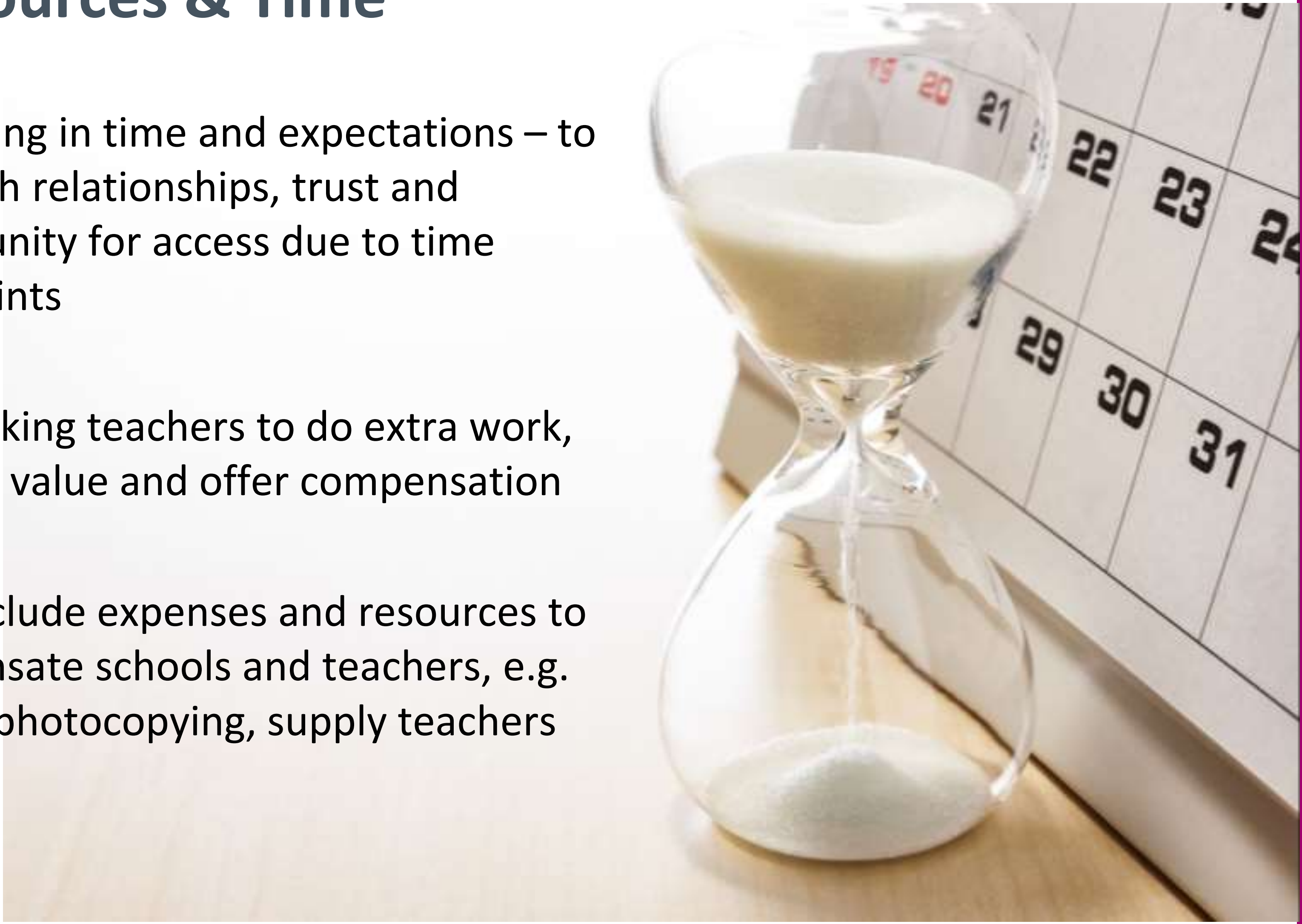
Providing beneficial outcomes for teachers – e.g. workshop, resources, PD

Resources & Time

Time building in time and expectations – to establish relationships, trust and opportunity for access due to time constraints

Ethically asking teachers to do extra work, need to value and offer compensation

Funds – Include expenses and resources to compensate schools and teachers, e.g. cost of photocopying, supply teachers



Collating the Corpus: Instructions for teachers…

1. Erase the students name and any identifying information and give each student in your class a case number.

2. Keep a record of the student’s case number given to student, as hopefully you will submit more texts for your class, and you can reuse the same number for each student.

3. Write the student number on the text you give to the research team.

4. Include as much information for each student as you can in the form below. If you do not have the information, please indicate with NK (Not Known).

5. A folder has been set up on one drive for you to upload the texts you wish to share.

6. You will be given a link to a folder that is on a GDPR data storage safe space. The folder should have your initials on it. Upload a folder (with a date as the title) containing a set of texts and a task info cover sheet.

BAWESS Corpus Meta

Student Number	Exam type: A-level/ GCSE/ iGCSE/ IB	Grade	Marke
Case 1			
Case 2			
Case 3			
Case 4			
Case 5			
Case 6			

Exam type: A-level/ GCSE/ iGCSE/ IB

Grade

Marked out of

Exam board

Genre/ Purpose of text

Condition / context of written work

Gender of student

EAL student

Special/ additional needs student

Number of years in a British (type) education system

Length of experience in the UK

Key barriers: Collecting Consent

"It’s a very painful thing, sending out, chasing, explaining, collating".

Offer assistance, e.g. send an RA to support, send drafted emails, etc.

Make participant information sheets & consent forms digital

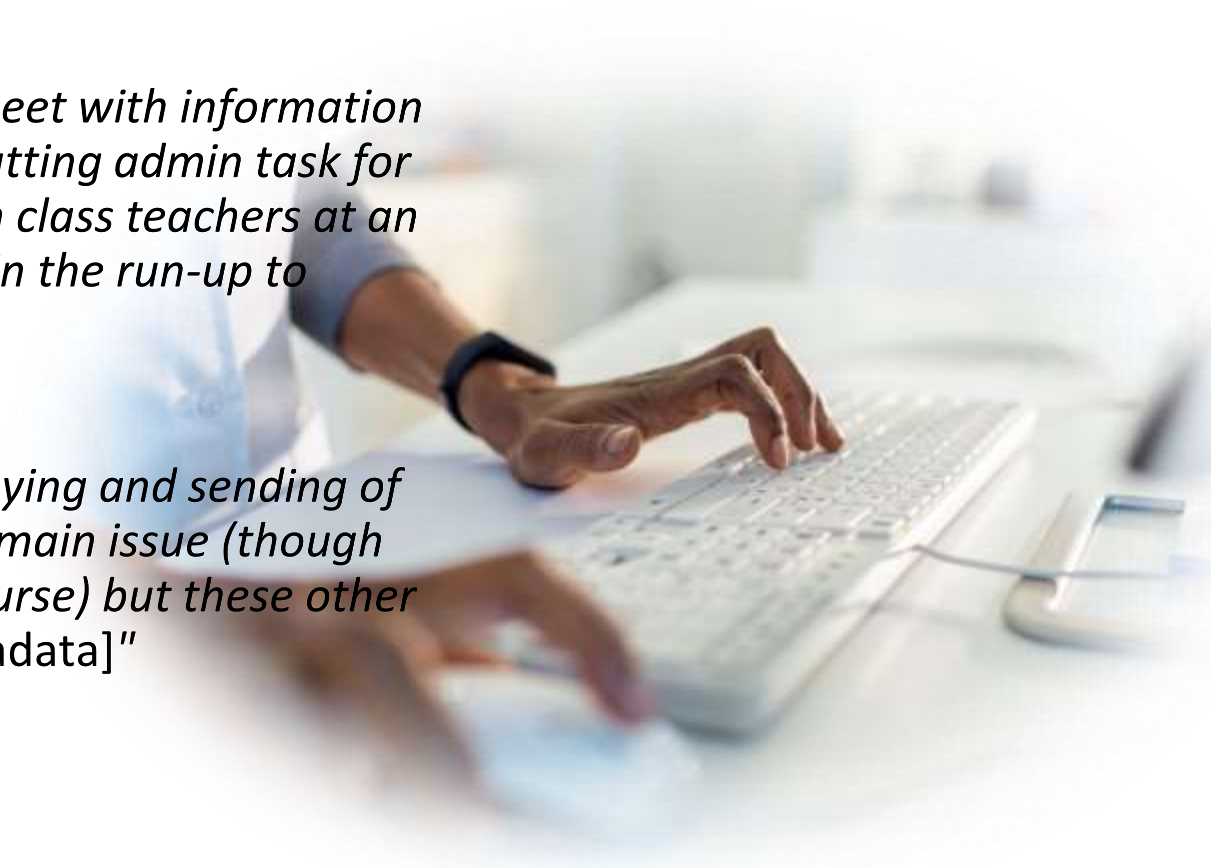
Emphasise value & benefits of involvement

Recording the metadata: Teachers' responses

"not that onerous as the data is in school, but depends on how and where it’s kept"

"repopulating a spreadsheet with information for each pupil is an off-putting admin task for teachers, especially exam class teachers at an excruciatingly busy time in the run-up to exams"

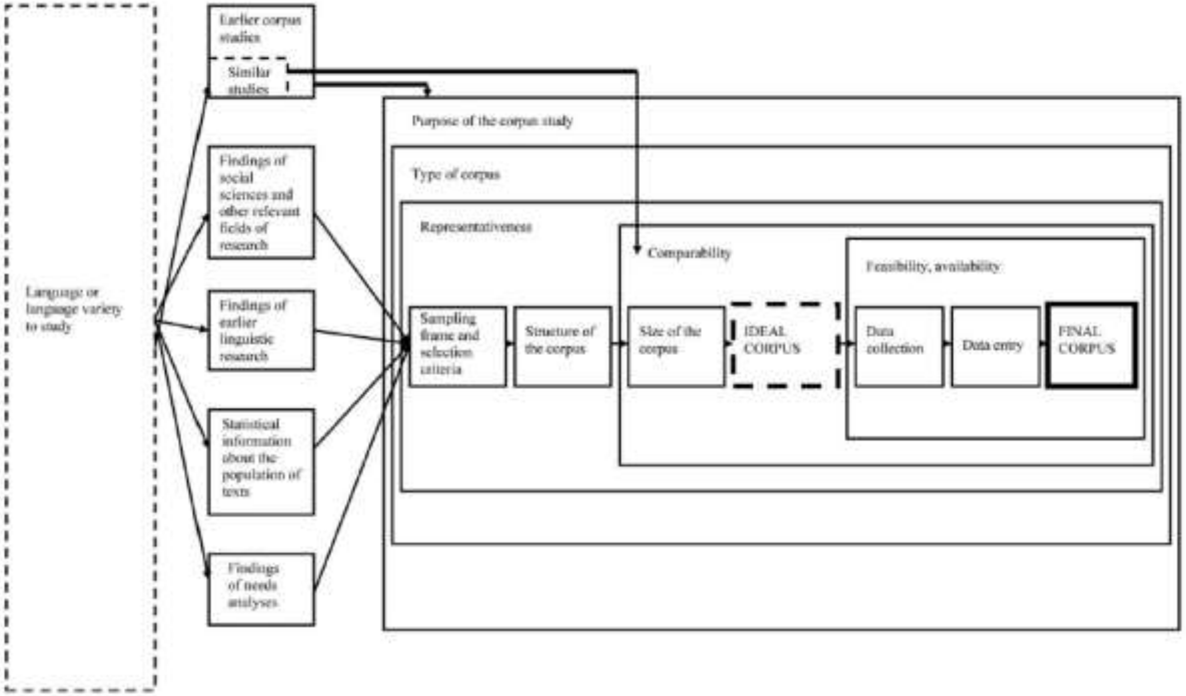
"It’s probably not the copying and sending of pupil’s writing that’s the main issue (though this also takes time of course) but these other things [consent and metadata]"



Methodological issues: Corpus design and preparation

Corpus design

- Age
- Subjects
- Quality
- Genre


Language or language variety to study
Earlier corpus studies
Similar studies
Findings of social sciences and other relevant fields of research
Findings of earlier linguistic research
Statistical information about the population of texts
Findings of needs analyses
Purpose of the corpus study
Type of corpus
Representativeness
Comparability
Feasibility, availability
Sampling frame and selection criteria
Structure of the corpus
Size of the corpus
IDEAL CORPUS
Data collection
Data entry
FINAL CORPUS

ure 30.2 Framework for corpus building

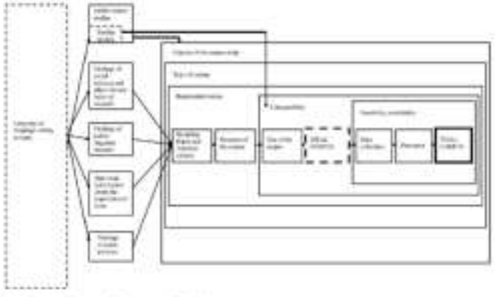
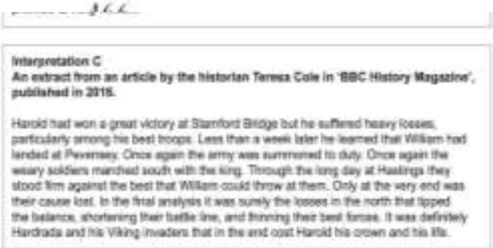


Figure 30.2 Framework for corpus building



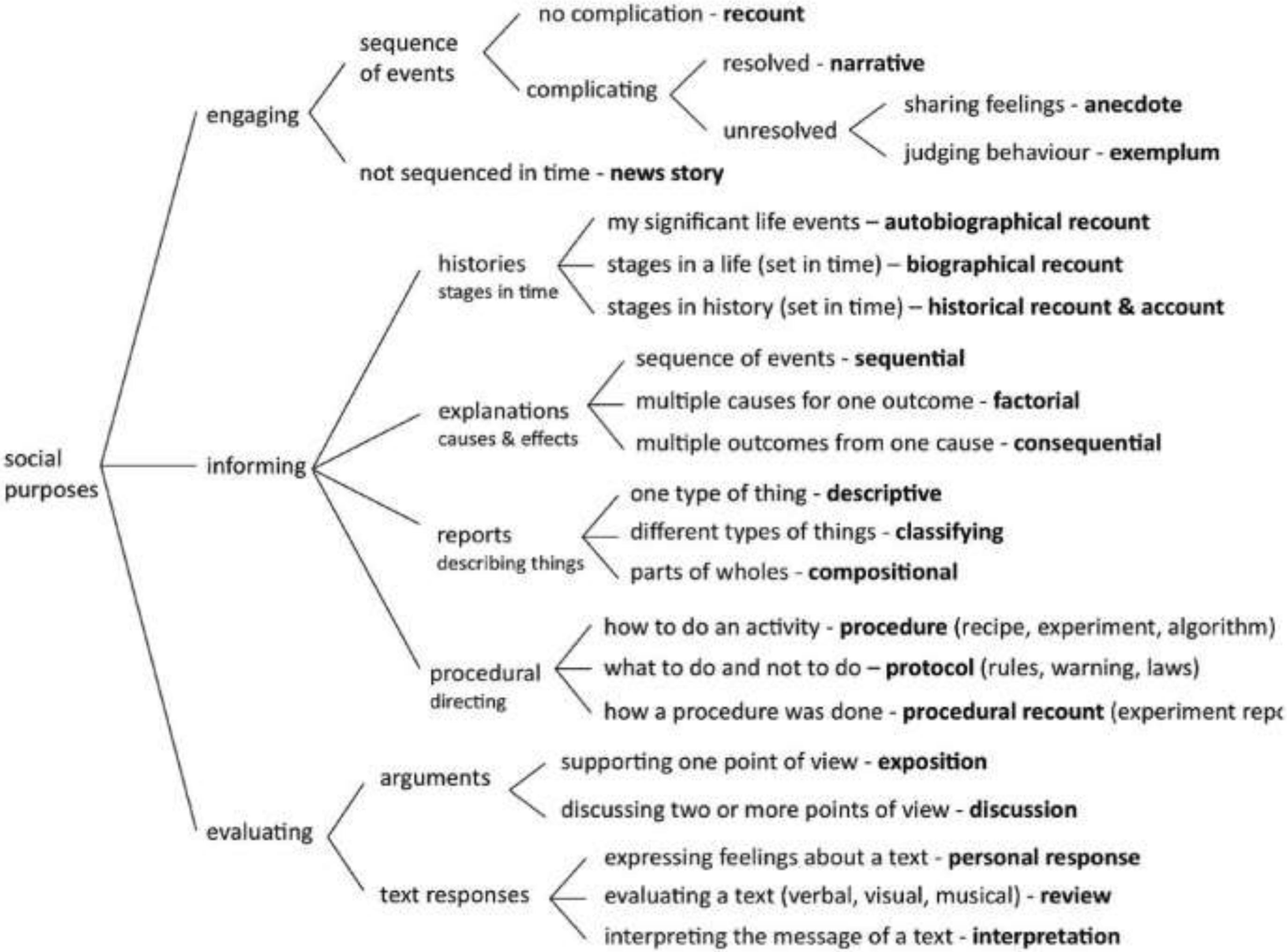
Interpretation C
An extract from an article by the historian Teresa Cole in 'BBC History Magazine', published in 2016.

Harold had won a great victory at Stamford Bridge but he suffered heavy losses, particularly among his best troops. Less than a week later he learned that William had landed at Pevensey. Once again the army was summoned to duty. Once again the weary soldiers marched south with the king. Through the long day at Hastings they stood firm against the best that William could throw at them. Only at the very end was their cause lost. In the final analysis it was surely the losses in the north that tipped the balance, shortening their battle line, and thinning their best forces. It was definitely Hardrada and his Viking invaders that in the end cost Harold his crown and his life.

(Jablonkai, 2022)

Genres in school

(Martin & Rose 2012: 110)



- social purposes
  - engaging
    - sequence of events
      - no complication - **recount**
      - complicating
        - resolved - **narrative**
        - unresolved
          - sharing feelings - **anecdote**
          - judging behaviour - **exemplum**
    - not sequenced in time - **news story**
  - informing
    - histories stages in time
      - my significant life events – **autobiographical recount**
      - stages in a life (set in time) – **biographical recount**
      - stages in history (set in time) – **historical recount & account**
    - explanations causes & effects
      - sequence of events - **sequential**
      - multiple causes for one outcome - **factorial**
      - multiple outcomes from one cause - **consequential**
    - reports describing things
      - one type of thing - **descriptive**
      - different types of things - **classifying**
      - parts of wholes - **compositional**
    - procedural directing
      - how to do an activity - **procedure** (recipe, experiment, algorithm)
      - what to do and not to do – **protocol** (rules, warning, laws)
      - how a procedure was done - **procedural recount** (experiment rep
  - evaluating
    - arguments
      - supporting one point of view - **exposition**
      - discussing two or more points of view - **discussion**
    - text responses
      - expressing feelings about a text - **personal response**
      - evaluating a text (verbal, visual, musical) - **review**
      - interpreting the message of a text - **interpretation**

School Genres

LITERARY TEXT TYPES:

LITERARY	TEXT TYPES EXAMPLES OF LITERARY TEXT FORMS
Narrative	novel, short story, myth, legend, science fiction, fantasy, fable, cartoon, stage play, film script, television script, radio script, role play
Poetry	sonnet, haiku, lyric verse, song, limerick, jingle, epic, ballad

FACTUAL TEXT TYPES:

Genre	TEXT TYPES EXAMPLES OF FACTUAL TEXT FORMS (Modes)
Report	reference book, documentary, guidebook, experimental report, group presentation
Recount	journal, diary, newspaper article, historical recount, letter, log, timeline
Procedure	instruction, recipe, directions
Exposition	advertisement, lecture, editorial, letter to the editor, speech, newspaper article, magazine article
Explanation	scientific writing, spoken presentation
Description	observation, speech, analysis
Response	book review, film review, restaurant review, personal response
Discussion	debate, conversation, talkback radio

From: http://australiancurriculumresourcesf-6.wikispaces.com/Literacy+Resources

Methodological issues

Corpus preparation

Principles for transcription (Smith et al., 1998)

1 Fidelity to the original

2 Consistency

3 Research deadlines

4 Research goals

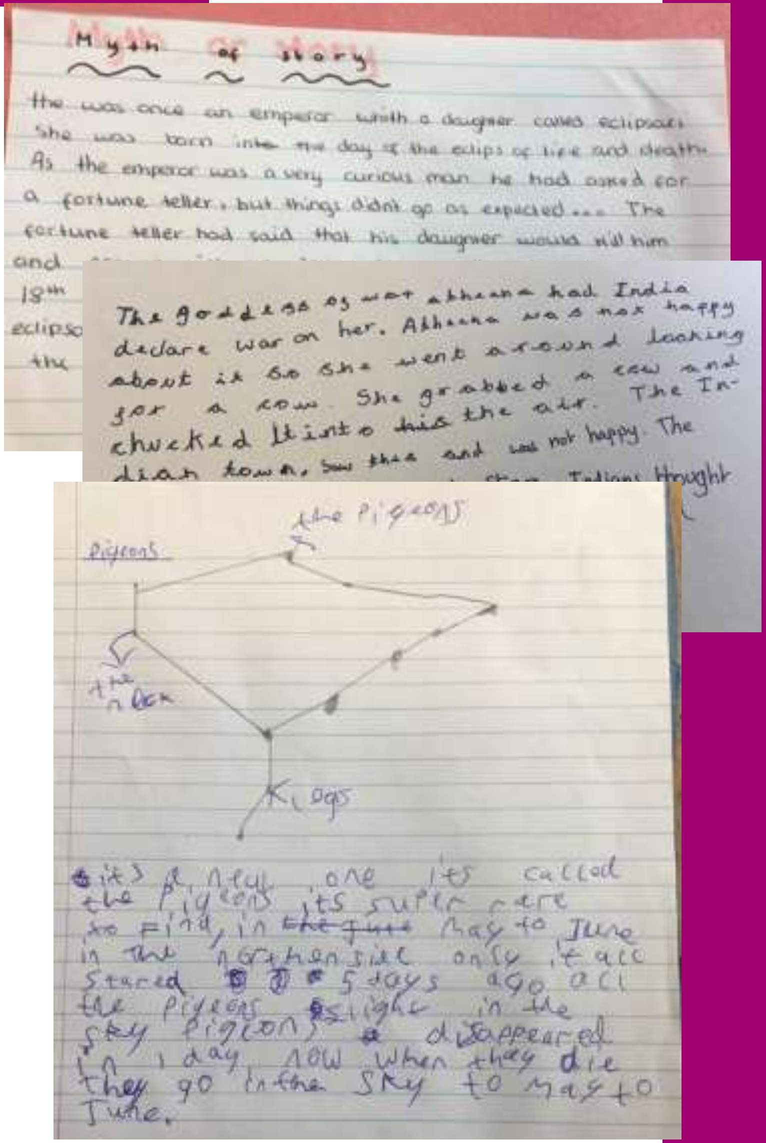
5 Observation of standard practices

Two-step transcription (Durrant & Benchley, 2018)

Authenticity, multi-modality (Gold et al., 2023)

Error tagging (e.g. Thewissen, 2013)

Two versions of the corpus



Myth of story

the was once an emperor whith a daugher calles eclipsoa.
She was born into the day of the eclips of life and death.
As the emperor was a very curious man he had asked for
a fortune teller, but things didn't go as expected ... The
fortune teller had said that his daugher would kill him
and [illegible]
18th
eclipso
the

The goddess of war athena had India
declare war on her. Athena was not happy
about it so she went around looking
for a cow. She grabbed a cow and
chucked it into ~~his~~ the air. The In-
dian town, saw the and was not happy. The
[illegible] Indians thought

its a new one its called
the Pigeons its super rare
to find, in ~~the june~~ May to June
in the northen side only it all
stared 5 days ago all
the pigeons slight in the
sky pigeons disappeared
in 1 day now when they die
they go in the sky to May to
June.

Conclusion

Focus on microethics and ethics in key moments in the research process

Access to text writers

Ethical considerations around working with young persons

Metadata

Use, collection, and preparation of texts

Access to corpus for research purposes

Access to corpus for instructional purposes

Methodology-specific ethical guidance

Ethics of access

Ethics of participation

Ethics of interpretation/ representation

Ethics of dissemination/intervention

Methodological considerations

handwritten texts, orthographic variation and multimodal elements in texts

Thank you

Dr Reka Jablonkai & Professor Gail Forey
rirj20@bath.ac.uk gf370@bath.ac.uk

References

Academy of Social Sciences (2015). Developing generic principles for social science research. https://acss.org.uk/wp-content/uploads//Developing-Generic-Ethics-Principles-for-Social-Science-Research-2015.pdf

BAAL (1994, 2006,2016, 2021). Recommendations on Good Practice in Applied Linguistics. https://www.baal.org.uk/wp-content/uploads/2021/03/BAAL-Good-Practice-Guidelines-2021.pdf

De Costa, P. I. (Ed.). (2015). *Ethics in Applied Linguistics research*. Routledge. https://doi.org/10.4324/9781315816937

De Costa, P., Sterling, S., Lee, J., Li, W., & Rawal, H. (2021). Research tasks on ethics in applied linguistics. *Language Teaching, 54*(1), 58-70.

De Costa, P. I. (2024). What's ethics got to do with applied linguistics? Revisiting past, considering the present, and being optimistic about the future of our field. *Research Methods in Applied Linguistics, 1*(3). https://doi.org/10.1016/j.rmal.2024.100103

Bell, P., Collins, L. & Marsden, E. (2022). Managing oral and written data from an ESL corpus from Canadian secondary school students in a compulsory, school-based ESL program. In Berez-Kroeker A., McDonnell, B., Koller, E. & Collister, L., (Eds.) *The Open Handbook of Linguistic Data Management. MIT Press , MIT, pp. 401-409.*

Gold, C., Laarman-Quante, R., & Zesch, T. (July, 2023). Preserving the authenticity of handwritten learner language: Annotation guidelines for creating transcripts retaining orthographic features. *Proceedings of the Annual Meeting of the Associatin for Computational Linguistics*, (CAWL 2023), pp. 14–21.

Elliot, M., Mackey, E., & Kieran O'Hara (2020). The Anonymisation Decision-Making Framework: European practitioners' Guide. UKAN University of Manchester. https://eprints.soton.ac.uk/445373/%0Ahttps://eprints.soton.ac.uk/445373/1/adf_2nd_edition_1.pdf

Hamid, M., & Crosthwaite, P. (2022). Developmental corpus insights into the writing life of a primary school child in Australia. Australian Review of Applied Linguistics, https://doi.org/10.1075/aral.21062.ham

Jablonkai, R. R. (2022). Building corpora for ELT. In R. R. Jablonkai & E. Csomay (Eds). *The Routledge Handbook of Corpora and English Language Teaching and Learning*. Routledge.

Kubanyiova, M. (2008). Rethinking research ethics in contemporary applied linguistics: the tension between macroethical and microethical perspectives in situated research. *Modern Language Journal, 92*(4): 503 -18.

Leedham, M., Lillis, T.,& Twiner, A. (2021). Creating a corpus of sensitive and hard-to-access texts: Methodological challenges and ethical concerns in the building of the WiSP Corpus. Applied Corpus Linguistics, 1, 100011. https://doi.org/10.1016/j.acorp.2021.100011

Lillis, T., Twiner, A., Balkow, M., Lucas, G., Smith, M., & Leedham, M. (2023). Reflections on the procedural and practical ethics in researching professional social work writing. *Journal of Applied Linguistics and Professional Practice*. https://doi.org/10.1558/jalpp.20014

Llaurado, A., Marti, A. & Tolchinsky, L. (2012). Corpus CesCa Compiling a corpus of written Catalan produced by school children. *International Journal of Corpus Linguistics 17*(3) (2012), 428–441. doi 10.1075/ijcl.17.3.06lla

McEnery, T. & Brookes, G. (2022). Building a written corpus: what are the basics? In A. O'Keeffe & M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics,* (pp. 35-47). Routledge.

McEnery, T. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.

Rose, D. & Martin. J.R. 2012. *Learning to Write, Reading to Learn: Genre, Knowledge and Pedagogy in the Sydney School.* London: Equinox**.**

Smith, N., McEnery, T. & Ivanic, R. (1998). Issuea in transcribing a corpus of children's handwritten projects. *Literary and Linguistic Computing, 13*(4), 217-225.

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *Modern Language Journal, 97*(1), 77-101.

Tilley, L. & Woodthorpe, K. (2011). Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research 11*(2) 197 –212.