Creating a corpus of social workers’ writing: Methodological challenges, representational issues and analytical concerns

Conference or Workshop Item

How to cite:


For guidance on citations see FAQs.

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
Creating a corpus of social workers’ writing: Methodological challenges, representational issues and analytical concerns

Maria Leedham, Theresa Lillis and Alison Twiner
The Open University, UK
Outline

1) Overview of WiSP research project

2) Building the WiSP corpus:
   - methodological challenges
   - representational issues
   - analytical concerns

3) Some initial findings
The WiSP research project

- 3 year, ESRC-funded research project (Oct 2015- Oct 2018)
- The first national research project on writing in professional social work practice

Research Questions (for the whole project)
1. What are the institutional writing demands in contemporary social work?
2. What are the writing practices and perspectives of professional social workers?
3. What are the challenges faced and solutions found?
4. How are writing demands and practices shaping the nature of professional social work?

Builds on existing work
- Writing/literacies research (e.g. Barton and Hamilton; Barton and Papen; Lillis; Street),
- Everyday workplace literacies research (e.g. Brandt; Tusting; Smith)
- Management/recording systems and practices in social work (e.g. White et al; Huuskonen and Vakkari; Taylor)
- Professional discourse studies - oral (e.g. Roberts & Sarangi; Sarangi)
- Research on writing in social work (e.g. Pare; Rai and Lillis)
Why is this project important?

The production and use of written texts (often referred to as *paperwork, recording, inputting or documenting*) is a **high stakes activity** in professional social work.

Writing…

- plays a central role in all decisions about services for people
- is used to evaluate social workers’ professional competence
- is often criticised – in public reviews and media reporting of high profile cases

Yet there is little systematic research on contemporary writing/recording demands, genres and practices….
Ethnography as our overarching epistemological orientation

There is no way in which language can be ‘context-less’ in this anthropological tradition in ethnography. To language, there is always a particular function, a concrete shape, a specific mode of operation, and an identifiable set of relations between singular acts of language and wider patterns of resources and their functions. (Blommaert, 2006, p. 4)

- Context-sensitive data collection methods and analysis in order to respond to research questions...
- A range of data...including interviews, observations, texts (case notes, emails, notes..), documentary data.

- **Corpus data is one of multiple data sets** within this ethnographic epistemological orientation.
- Aim of building the corpus is to assist in investigating SWs’ written texts

- **Advisory panel of stakeholders:** social workers, SW managers, HEA, professional bodies e.g. HCPC
Data collection almost complete

- 5 agencies participating
- 70 social worker interviews
- 10 weeks of researcher observations
- 481 days of social worker writing activity logs
- 4,600 texts collected and anonymised
- 1 mw corpus

Still to do… screen capture of social worker at-desk writing
Hello [FA1],

Please ignore the previous blank one!

Please see below – relates to PIN [NUMBER]. I agreed with [EST1] that their support would end as of 22/1/2016, thereby ending the Direct Payment arrangement too.

Is this enough information for you to end the service as it is? There will then be further charges on the way as [COUNCIL2] are assessing and commissioning service, which [COUNCIL2] will be paying for under their s.117 responsibilities. Is it somehow possible to keep the Direct Payment 'open' pending this?

Hope this makes sense.

Thanks again,

[SW005] (Social Worker, [LOCATION])

‘[OA1] carers not turning up to provide meals support, twice this week, [SU008a] rang care line who alerted [OA1] who did not respond, [SU008a] rang [OA1] 9.30am this morning they said 'sorry' gave no other explanation, they sent a carer 10.30am instead of 7.30am. [SU008a] managed to make himself a slice of toast this morning and stated he can't stand too long to prepare his meals due to a weak left leg [h]as a result of a weak hip.

Contracts monitoring form to [be] completed, [SU008a] will have in - house carers if [DAUGHTER] agrees.

T/c message to [DSO] for advice on their capacity, t/c to [DAUGHTER] messages left X 2.

Spoke to [DSO] who rang to say she can provide the support hours and will get back to me with a start date.


I advised that [SU008a] has 4 calls today with the names of 4 different carers, he usually has [CARER1] and [CARER2] which is the plan for the rest of the week.
Methodological challenge 1: Gaining access to sensitive texts...
Methodological challenge 1: Gaining access to sensitive texts...

1. **Institutional** - formal research and governance procedures to seek permission from social services to carry out research.
   - LA 1 - (previous formal contact) level 1 secured at 6 and 12 months, insider contact ensured 2 and 3, **approved, strong**.
   - LA 2 - (previous formal contact) level 1 secured at 8 months, level 3 secured but failed because rejected at level 2, **rejected**.
   - LA 3 - (previous formal contact) additional legal agreement required around data protection, level 1 secured at 9 months, insider contact ensured 2 and 3, **approved, strong**.
   - LA 4 - (no previous formal contact) level 1 secured at 3 months, no support at level 2, slow development, **approved, weak**.
   - LA 5 - (no previous formal contact) and eventually rejected, renegotiated lower level of involvement secured at 12 months, level 3 some support, **approved, weak**.
   - LA 6 - (no previous formal contact) after several months we negotiated access to low level involvement, **approved, weak**.

2. **Operational** - permission from service managers to approach social workers, depends on teams having ‘capacity’ and assurance that research will not disrupt practice.
   - LA 2 - (previous formal contact) level 1 secured at 8 months, level 3 secured but failed because rejected at level 2, **rejected**.
   - LA 3 - (previous formal contact) additional legal agreement required around data protection, level 1 secured at 9 months, insider contact ensured 2 and 3, **approved, strong**.
   - LA 4 - (no previous formal contact) level 1 secured at 3 months, no support at level 2, slow development, **approved, weak**.
   - LA 5 - (no previous formal contact) and eventually rejected, renegotiated lower level of involvement secured at 12 months, level 3 some support, **approved, weak**.
   - LA 6 - (no previous formal contact) after several months we negotiated access to low level involvement, **approved, weak**.

3. **Individual** - agreement from individual social workers to take part (in different ways and levels) – some with payment.
Ethical issues....

Initial ethical approval from ESRC, OU, Local Authorities.... Ethics as an ongoing consideration for the ‘virtuous researcher’ (Oates)

But - Who should we ask for consent in using texts?

- Standard ethical procedure is to ask for consent of research participants. But who should we ask? - The Local Authority responsible for assigning the case? The social worker and writer of the text? Or the service user/client and focus of the text?

What we did

- gained consent from each participating Local Authority to access texts written by social workers – with proviso that texts were anonymised in-house and before we had access
- gained permission from the social worker-writers
- did *not* ask service users for their consent to use texts
- CF medical research (Mann et al, 2016)
Methodological challenge 2: Anonymising the data

- Social workers write sensitive reports, often about vulnerable people, and often including personal or sensitive information.

What is personal or sensitive data?
- *Personal data* relates to data that can identify a service user or other individual such as their home address or their date of birth.
- *Sensitive personal data* is information that has a greater risk of damage to the individual if it is misused or mishandled, such as criminal convictions, physical or mental conditions (Elliot et al, 2016).

How was anonymization carried out?
- Local Authority carried out the initial *redaction* of personal identifying features (name, GP name, etc) on site.
- Asked redactors to replace personal details with *codes* e.g. [TEL], [CITY], [SU], [HEALTH], [SCHOOL], [CITY]
- Also… [SUD], [MUM EX-PARTNER], [BIRTH_MUM], [CHILD3]
Action - Ongoing Record

Phone call to [SCHOOL] school([TEL]) to arrange a direct work session to see [SU]. Visit arranged for the 2\textsuperscript{nd} January at 11:30am.
Action - Ongoing Record

[SW201] Phone call to [MOTHERS-EX-PARTNER] (father to [HALF-SIBLING_2]). I spoke to [MOTHERS-EX-PARTNER] about the issues [SU-MOTHER] had raised regarding her contact with [HALF-SIBLING_2]. [MOTHERS-EX-PARTNER] said that [SU-MOTHER] can ring and speak to [HALF-SIBLING_2] anytime she wants. [MOTHERS-EX-PARTNER] said that issues was more that [HALF-SIBLING_2] has little interest in speaking to anyone on the phone as he would rather play and his interest has to be sparked on the phone. [MOTHERS-EX-PARTNER] said that he has not spoken to [SU-MOTHER] about contact so has not said that [HALF-SIBLING_2] can't go and stay and his view was that [SU MOTHER] is playing the martyr. [MOTHERS-EX-PARTNER] was clear that he wants [HALF-SIBLING_2] to have a relationship with [SU-MOTHER], and that she is more than welcome to pick him up and take him out for tea, but he suspects she will not do this as it requires her to make an effort.
Case note 3

Chronology

[DATE] - Incident with knife - No information relating to this found on [IT_SYSTEM].

[DATE] - Telephone call to CRU re. Concerns raised. Contact made with dad. [PERSON] had stopped taking meds. Dad states he does not require support at this time. NFA.

[DATE] - Referral to MAT. Mum taken significant overdose 4 x 28 [DRUG] Admitted to ITU.

[DATE] - Referral to MAT. Mum taken overdose 14 - 16 [DRUG] Admitted to hospital

[DATE] - Call from [HOSPITAL]. Mum transferred to CDU after taking an OD of anti-depressants and drinking a bottle of vodka. Previous attendance at [LOCATION] Hospital.

[DATE] - Child referral.


[DATE] - [SU] exhibited sexualised behaviour by pulling a girls trousers down and asking her to do the same. Witness by Midday Supervisor in school.

[DATE] - Referral from school. Information that mum is having sexual relations with her brother. He has moved into the house.
Methodological challenge 3: Archiving the corpus

Initial anonymization

Filename: SW207_casenote_120315_3
I advised that [SU207a] has 4 calls today with the names of 4 different carers, he usually has [CARER1] and [CARER2] which is the plan for the rest of the week, [PERSON3] apologised and stated …

Final anonymization

Filename: WISP1679
I advised that [SU] has 4 calls today with the names of 4 different carers, he usually has [CARER] and [CARER] which is the plan for the rest of the week, [PERSON] apologised and stated …

• => Standalone texts rather than part of a trajectory
• Will restrict user access to 'permission' level in UKDA
• Two corpora: team and archived versions

Archiving the corpus

• Checked anonymization using Wmatrix semtagging and through reading the corpus
• Added further codes (e.g. for rare medical conditions).
• Removed all dates [DATE]
• Scaled down the metadata associated with texts
• Delinked texts from their writer [SW] not [SW207]
• Changed [CHILD2] to [CHILD] etc
Representational issues

• How do we attempt to represent the writing of social workers?
• Sampling model? – across LAs, domains, text types, texts,
• Have a variety of social worker writers by age/ experience etc?
• All texts are not equal (update casenote vs critical incident casenote vs internal email vs court report)
Texts for a single case

Written documents related to ‘children at risk’ case during week of observation

- Emails
- Sibling 1 assessment
- Sibling 2 assessment
- Sibling 3 assessment
- Parental assessment - mother
- Parental assessment - father
- Genogram update
- Handwritten notes of Personal Educational Plan meeting x 2
- Handwritten notes of ‘viability assessment’- of another family member taking care of one child
- Handwritten notes on parent-child contact visits observed
- Handwritten notes of statutory visit to child
Representational issues

• How do we attempt to represent the writing of social workers?
• Sampling model? – across LAs, domains, text types, texts,
• Have a variety of social worker writers by age/ experience etc?
• All texts are not equal (update casenote vs critical incident casenote vs internal email vs court report)
• Tag duplicates as don’t want to over-represent any writer?….. But what is a duplicate? (non-SW writing? e.g. boilerplate disclaimers, form questions, VS. repeated SW language e.g.formulaic language, emails within casenotes, multiple assessments in one family, repeating action points….)

But…

• Practical issues of access, time, money & goodwill

So… take what you can get

• And describe the corpus composition for future users
• Small corpus – 1 mw after deduping (from 1.2mw)
• NB Delinking of datasets and of writer from text makes the corpus less useful to other researchers
The WiSP corpus
(by wordcount)

- assessment reports (436,908) 44%
- casenotes (407,803) 41%
- emails (92,955) 9%
- other (65,423) 6%
Analytical issues… and some findings

- Wmatrix used to extract KW and MWUs (Rayson, 2008)
- BE06 as ref corpus
- Filter on frequency: minimum 50
- Filter on stats: Bayes 2+
- Filter on effect size: %DIFF
- Candidate key items halted at break in effect size (Gabrielatos, 2018)
- CKIs checked to ensure in min. 23 texts and by min. 5 SW writers plus across all 3 SW domains
- CKIs checked to exclude anonymization codes – [SU] as most key
- Thematically categorized – iterative process within the team of 3, assigned multiple categories if 30%+
- Asked advisory panel for insights around KW

<table>
<thead>
<tr>
<th>No</th>
<th>Category</th>
<th>Example key items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SW roles and activities</td>
<td>care*, LAC, proceedings, safeguarding,</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Our example:</strong> They shared all evening meals together.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= evaluative language?</td>
</tr>
<tr>
<td>2</td>
<td>Family &amp; institutionalised care</td>
<td>adopters, birth_mother, care*, contact*, foster, unborn,</td>
</tr>
<tr>
<td>3</td>
<td>Family &amp; institutionalised care</td>
<td>processing, assault, attends, bail, belongings, bruising, home_environment, meals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>residing, toilet, unwell, utilise,</td>
</tr>
<tr>
<td>4</td>
<td>Primarily describing a situation</td>
<td>accessing, assault, attends, bail, belongings, bruising, home_environment, meals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>residing, toilet, unwell, utilise,</td>
</tr>
<tr>
<td>5</td>
<td>Primarily evaluating</td>
<td>abusive, appropriately, behaviours, can’t, concerns, due_to, enjoys, exploitation,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>misuse routines, struggles,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[PERSON] continues to misuse heroin.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Emphasises illegal drug use/excessive alcohol use</td>
</tr>
<tr>
<td></td>
<td></td>
<td>She needs to have better routines and ensure…</td>
</tr>
<tr>
<td></td>
<td></td>
<td>= evaluative within SW domain of children</td>
</tr>
</tbody>
</table>
## Comparing text types

<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>Case notes</th>
<th>Emails</th>
<th>Assessment reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Primarily describing a situation</td>
<td>arrived</td>
<td>finance</td>
<td>achieve</td>
</tr>
<tr>
<td></td>
<td></td>
<td>asleep</td>
<td>allowance</td>
<td>appearance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>bail</td>
<td>funded</td>
<td>at_risk</td>
</tr>
<tr>
<td>5</td>
<td>Primarily evaluating</td>
<td>attempted</td>
<td></td>
<td>alleged</td>
</tr>
<tr>
<td></td>
<td></td>
<td>shouted</td>
<td></td>
<td>befriends</td>
</tr>
<tr>
<td></td>
<td></td>
<td>worried</td>
<td></td>
<td>illicit</td>
</tr>
<tr>
<td>6</td>
<td>Communication</td>
<td>advised</td>
<td>attached</td>
<td>caller</td>
</tr>
<tr>
<td></td>
<td></td>
<td>discussed</td>
<td>catch_up</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>suggested</td>
<td>copies</td>
<td></td>
</tr>
</tbody>
</table>

- Is it possible to separate ‘describing’ and ‘evaluating’?
- Is all SW writing evaluative?
The example of *asleep* – evaluative?

- [CHILD] looked comfortable with [MUM] and he fell *asleep* on her. He appeared content in the home and [MUM] was attentive to his needs. She fed him during the visit. (visit to toddler and parent)

- She is not keeping him clean, there was no fresh food at home. His bed was not made and all rooms were very dirty, untidy and smelly. [CHILD] said he sometimes slept on the sofa with Mum and fell *asleep* watching TV with her. The home conditions and parenting are not good enough at present. (statutory visit to a 10 year old child)

- [CHILD] stated he started a paper round to get extra money for himself however [CHILD] would fall *asleep* at school.

- We knocked at the door several times before we got an answer. [DAD] then came to the door and appeared very sleepy as though he had just been woken up. [CHILD] was wondering about and appeared happy. [DAD] invited us in. I asked [DAD] if he had been *asleep* and he stated he was very tired from doing his night shift, he stated he hadn't been *asleep* but his eyes had been shutting but he could still hear everything.

- [SU] still unsteady mobilising and spending a lot of time *asleep*. (casenote, adults)
Comparing text types

<table>
<thead>
<tr>
<th>No.</th>
<th>Category</th>
<th>Key items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case notes</td>
<td>Emails</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Told [CHILD] to apologise, shouted "for fucks sake...."
2. her with not going on holiday again, shouted 'no!' took her by the hand
3. Do you want to apologise for your behaviour?' shouted 'No!' then said
4. shouted from the lounge 'mum' shouted back to him 'if you want me come
5. go and see if [PERSON] wants to play?' shouted 'No' slammed the cupboard doors

[PERSON] stated that she is not sure how [PERSON] ([SU]'s son) will feel about this. She asked "do you think that they will think I am awful for not having him back". I advised that I have spoken with [PERSON] (Son) and he was open to the possibility of [SU] remaining in residential care. I advised [PERSON] that she has to be honest how she feels as [SU] could only return home if she is able and willing to care for him.

- Report in past tense on arrangements
- Different reporting of SW and SU talk (advised, discussed, agreed, shouted) – quotations to provide evidence

- Request notes
- Suggested
- Attached
- Copies
- Caller

- Look externally to other professionals and services ([surgery, housing])
- Use SU’s language ([nan, wife])
- Contain a greater specificity of times and dates for recent or forthcoming events (11am, last_night)
Conclusions aka *What I’ve learned*…

- Don’t underestimate the issues involved in compiling a corpus of sensitive texts!
- Stakeholders and participants can offer valuable insights
- Important to try out ways of slicing the data to ensure findings are robust
- Need to consider variables within the data
- Don’t over-promise what you can archive…
References


Thank you

@OUWISP #OUWISP

www.writinginsocialwork.com