



OPEN ACCESS

EDITED AND REVIEWED BY
Mila Vulchanova,
NTNU, Norway

*CORRESPONDENCE
Helen Fraser
✉ helen.fraser@unimelb.edu.au

RECEIVED 15 April 2024
ACCEPTED 22 April 2024
PUBLISHED 08 May 2024

CITATION
Fraser H, Haworth K, Deamer F, Loakes D,
Richardson E and Komter M (2024) Editorial:
Capturing talk: the institutional practices
surrounding the transcription of spoken
language. *Front. Commun.* 9:1417465.
doi: 10.3389/fcomm.2024.1417465

COPYRIGHT
© 2024 Fraser, Haworth, Deamer, Loakes,
Richardson and Komter. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Editorial: Capturing talk: the institutional practices surrounding the transcription of spoken language

Helen Fraser^{1*}, Kate Haworth², Felicity Deamer²,
Debbie Loakes¹, Emma Richardson^{2,3} and Martha Komter⁴

¹Research Hub for Language in Forensic Evidence, The University of Melbourne, Parkville, VIC, Australia, ²Aston Institute for Forensic Linguistics, Aston University, Birmingham, United Kingdom, ³Department of Communication and Media, School of Social Science and Humanities, Loughborough University, Loughborough, United Kingdom, ⁴Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, Netherlands

KEYWORDS

transcription, misconceptions about language and linguistics, language ideologies, forensic linguistics, forensic transcription, police interviews and interrogations, entextualization

Editorial on the Research Topic

[Capturing talk: the institutional practices surrounding the transcription of spoken language](#)

Transcripts are a ubiquitous feature of virtually all modern institutions, many of which would be unable to function without them. Nevertheless, transcription remains an under-researched subject—a situation that *Capturing talk: the institutional practices surrounding the transcription of spoken language* seeks to remedy.

The initial aim of this Research Topic was to expose and examine under-appreciated features of “entextualization” (the process of representing spoken language as written text). One of these features is the fact that a transcript can only ever be a representation of speech, not a copy—and thus can never represent speech exactly. Another feature, well-articulated by Sarangi (1998), is the unequal power over the process of transcription exercised by, on the one hand, the speakers whose voices are represented, and, on the other, by those controlling the transcription process.

Where Sarangi’s interest was mainly in health and social services institutions, the present Research Topic has a leaning toward legal institutions, where, arguably, these power inequalities are even more starkly contrasted—as demonstrated by the territory-defining volume (Heffer et al., 2013).

Four of the papers in this Research Topic deal with police interviews, providing insight into differing practices across jurisdictions and type of interview (e.g., whether with witnesses or suspects). Several papers examine the practice of converting an interview into a “statement,” written up by the officers who conduct the interviews. Beginning with interviews with witnesses in England and Wales (E&W), Milne et al. analyze a sample of such statements against transcripts produced by the researchers from an audio recording. The omissions, additions, distortions, and other errors in the police versions give cause for deep concern.

An extended study analyzing the creation of records of interviews with suspects in the Netherlands is recounted by [Komter](#), which, again, contrasts transcripts prepared by police interviewers, with the author's transcripts prepared from audio recordings. Again, many concerning limitations on the police transcripts are observed and analyzed. However, while her own transcripts are far more detailed, [Komter](#) acknowledges that she too is necessarily selective in what she chooses to represent, guided by the evolving research questions she seeks to investigate.

One practice [Komter](#) discusses is that of police records presenting an interview as a monolog, in the voice of the interviewee, rather than as the question-and-answer dialogue it actually was. This practice is also investigated by [Eerland and van Charldorp](#), again focusing on the Dutch context. These authors study how readers of the statements were influenced by three different styles of reporting (monolog, dialogue and narrative), with the troubling finding that the style of reporting affected perceptions of the statements' accuracy and comprehensibility.

In many jurisdictions, police interviews with suspects are routinely audio- or video-recorded. However, this does not signal the end of problems with the representation of these high-stakes interactions. The last of our interview papers is [Haworth et al.](#), which summarizes the key findings to date of an ongoing study of the transcription of electronic records of interviews with suspects in E&W. It demonstrates a range of problems with official police transcripts even when these ostensibly capture the dialogue "verbatim," and proposes that consistency, accuracy, and neutrality are the foundational features that should underpin any police interview transcript.

A second group of papers studies transcription in non-legal institutional settings. [Holder et al.](#) delves into two very large and highly structured organizations with serious security needs: NASA and the US Military. Both make extensive use of audio and video recordings capturing employees as they work—with transcripts produced either routinely, or on demand. The authors look into the two organizations' use of these transcripts, again comparing the official transcripts with their own transcripts of selected sections, using conversation analysis (CA) conventions.

[Park and Hepburn](#) also examine CA-style transcripts. Taking as an example Rachel Mitchell's interview of US Supreme Court nominee Brett Kavanaugh about his alleged historical sexual misconduct, these authors compare the information retrievable from a richly detailed Jeffersonian transcript with an orthographic transcript that "wipes out" or "skates over" crucial aspects of speech used by speakers and listeners in constructing the message expressed by the speech.

Another institutional use of transcripts covered in *Capturing Talk* concerns workers on the assembly line of a small factory in Sweden. [Carlsson and Harari](#) report an observation-and-interview study of the instruction manuals created by the workers. While they find much to commend in the retention of power by the creators and users of the manuals, the authors observe room for improvement in the "information design" of the texts, recommending that consultation of linguistics experts could offer benefits.

[Voutilainen](#) showcases the high quality of transcripts produced as an official record of the complex and challenging multicultural discussions of wide-ranging Research Topics covered by the

parliament in Finland. His account demonstrates how much thought, research and work goes into managing all the factors that need to be considered to create transcripts of this standard.

In a return to the legal setting, a further group of papers examines transcripts of forensic audio, i.e., recordings of speech used as evidence in criminal trials. These are often of very poor quality, meaning that the transcript is intended not as a record of what was said, but as assistance to the court in determining what was said. Internationally, it is common for such transcripts to be provided by police investigating the case. While the courts recognize that police transcripts might contain errors, they rely on judges and/or juries being able to check the transcript against the audio. This ignores well-established research findings that the very act of checking a transcript can cause the listener to hear in line with the transcript, even if it is demonstrably false. For this reason, linguists sometimes recommend that, to ensure accuracy, transcripts should be produced by independent experts in transcription.

However, mere independence may not be enough, and [Love and Wright](#) point out some important caveats around this recommendation. They had eight trained transcribers produce transcripts of poor-quality forensic-like audio—finding huge divergences in the content of the transcripts (<3% of conversational turns were transcribed consistently by all eight participants). This demonstrates that transcribing poor-quality forensic audio needs not just expertise in linguistics, but a managed, evidence-based method.

Recently, a common response to any discussion of the difficulty of transcribing poor-quality audio has been: "Why not let AI do it?" [Loakes](#) investigates this suggestion, finding that, while modern automatic speech recognition (ASR) systems are extremely efficient at transcribing good-quality audio, their performance on poor-quality forensic-like audio is low. Even the best-performing system, Whisper, scored only around 50% accuracy, with others far lower.

[Harrington](#) also observed low scores for ASR transcripts of poor-quality forensic-like audio. Bridging two of the main areas considered in this Research Topic, she also trialed ASR on recordings of police interviews. The resulting transcripts, though not problem-free, score far higher than those of covert recordings, with errors easier to identify. [Harrington](#) makes innovative recommendations for how ASR could be used as a "first draft" interview transcript, to be refined via human transcribers.

Two papers consider the transcription and translation of forensic audio featuring languages other than English. [Gilbert and Heydon](#) look at translated transcripts of Vietnamese recordings used as evidence in a drug-related trial. They point out significant errors in the translations, but note that, unless the defense goes to the expense of hiring their own translator/interpreter, such errors are unlikely to be detected—and suggest that audio in languages other than English is often admitted with inadequately tested translations.

[Lai](#) presents results of a large national survey of the practices and concerns of translators and interpreters who undertake forensic casework across a wide range of languages. Here, too, results indicate a number of important deficiencies in current practice for translating forensic audio featuring languages other than English—and [Lai](#) makes valuable recommendations for improvement.

Finally, taking an authoritative overview of the key issues relevant to this Research Topic, Fraser provides a systematic review of interdisciplinary research on transcripts and transcription, and sets out a series of interacting factors that are known to affect a transcript's reliability. Using examples from a range of legal and academic situations, Fraser argues that, to ensure a transcript is suitable for its intended purpose, it is essential that all the factors be appropriately managed.

Taken as a whole, *Capturing Talk* amplifies two observations made in both Sarangi (1998) and Heffer et al. (2013), which, though not the exclusive focus of any individual paper, are highlighted throughout the Research Topic. First, the strong role that context inevitably plays in the interpretation of a transcript implies that "recontextualization" (using a transcript in a context other than the one it was created in) is likely to change its interpretation. Second, even the most expert linguistic analysis of transcripts produced by others is not itself a neutral or "objective" activity. However, this does not mean that such analysis must be "subjective" in any limiting sense. Rather it indicates a need for transcripts to be produced and analyzed by independent, context-aware experts able to devote appropriate attention to all relevant factors.

Most importantly, all contributions to *Capturing Talk* emphasize that transcription is far from the simple transduction of "sounds" into letters that it is often assumed to be by those who have not studied its intricacies. It is a highly complex and fascinating Research Topic worthy of taking its place as a dedicated field of research in its own right, particularly in view of the widespread misconceptions and unhelpful language ideologies that still beset the institutional practices surrounding the transcription of spoken language.

References

Heffer, C., Rock, F., and Conley, J. (2013). *Legal-Lay Communication: Textual Travels in the Law*. Oxford: Oxford University Press.

Author contributions

HF: Writing – original draft, Writing – review & editing. KH: Writing – review & editing. FD: Writing – review & editing. DL: Writing – review & editing. ER: Writing – review & editing. MK: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Sarangi, S. (1998). Rethinking recontextualization in professional discourse studies: an epilogue. *Text Talk* 18, 301–318. doi: 10.1515/text.1.1998.18.2.301