



A Framework for Deciding How to Create and Evaluate Transcripts for Forensic and Other Purposes

Helen Fraser*

Research Hub for Language in Forensic Evidence, School of Languages and Linguistics, The University of Melbourne, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:

Dominic Watt,
University of York, United Kingdom

Reviewed by:

Alison May,
University of Leeds, United Kingdom
James Tompkinson,
Aston University, United Kingdom

*Correspondence:

Helen Fraser
helen.fraser@unimelb.edu.au

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 17 March 2022

Accepted: 16 June 2022

Published: 22 July 2022

Citation:

Fraser H (2022) A Framework for
Deciding How to Create and Evaluate
Transcripts for Forensic and Other
Purposes. *Front. Commun.* 7:898410.
doi: 10.3389/fcomm.2022.898410

Transcripts are used successfully in many areas of contemporary society. However, some uses of transcripts show systemic problems, with significant negative consequences. The key to finding effective solutions in these areas is to determine which factors contribute most strongly to the problems – which may be different from those to which they are commonly ascribed. This systematic review offers a conceptual framework for understanding the nature of transcripts in general, and the factors that contribute to a transcript's reliability and suitability for purpose. It then demonstrates how the framework can explain the (mostly) successful use of transcripts in two domains: court proceedings and linguistics research. Next, it uses the framework to examine two problematic cases: transcripts of forensic audio used as evidence in criminal trials, and transcripts of police interviews with suspects. A crucial observation is that, while it is common, and understandable, to focus on the transcriber as the source of problems with transcripts, transcription is actually a complex process involving practitioners in multiple roles, of which the transcriber role is not always the most important. Solving problems thus requires coordination of a range of factors. The analysis ends with practical suggestions for how to seek solutions for both the problematic areas reviewed, with attention to the role that linguistic science needs to play. The conclusion amplifies recent calls to consolidate transcription as a dedicated field of study within linguistics.

Keywords: transcription, transcript reliability, forensic, legal, verbatim reporting, covert recordings, police interviews, linguistic analysis

1. INTRODUCTION

Transcripts are an essential part of our literate culture, providing a convenient and lasting record of otherwise ephemeral spoken language (Olson, 1994). Their ubiquity and familiarity make transcription seem like a simple and unproblematic process. However, it has many hidden complexities which not only cause problems, but make those problems hard to identify and solve.

The focus of the present paper is on transcripts used in legal contexts – specifically on transcripts of court proceedings, police interviews and covert recordings, as used in Australian and UK jurisdictions. As will be seen, while transcripts of court proceedings are mostly handled well (though with important exceptions), transcripts of interviews and covert recordings show systemic problems known to create a threat to justice (see Bucholtz, 2009; French and Fraser, 2018; Haworth, 2018).

Transcripts are also used in many branches of linguistic research, such as phonetics (e.g., Heselwood, 2013), language description (e.g., Himmelmann, 2018), conversation analysis (e.g., Hepburn and Bolden, 2012), discourse analysis (e.g., Edwards, 2008) – and indeed in studies of language used in the legal process (see Coulthard et al., 2020). However, with some notable exceptions (see Jenks, 2013), transcription is usually discussed in relation to specific branches of linguistic research, rather than as a general topic in its own right. This is unfortunate, as it means scholars may lack awareness of relevant issues from other branches, making it more difficult to determine the best solution for problems such as those mentioned above.

This systematic review aims to consolidate transcription as a dedicated field of research spanning multiple branches of linguistic science (cf. Fraser, 2020b). It starts by drawing together research findings about transcription, some of which, though well established, are subject to substantial misconceptions outside their own specialised areas. It then outlines a general framework for thinking about the stages involved in creating and using a transcript, and the factors that need to be managed at each stage to ensure a reliable product suitable for its purpose. Next it shows how consideration of the factors can help explain the successful use of transcripts in two very different contexts: court proceedings and linguistic research. Finally it uses the factors to identify the causes of systemic problems with transcripts of forensic audio and of police interviews, and to offer suggestions for effective solutions. A strong theme is that developing effective solutions for these serious problems requires the linguistic sciences not just to apply existing knowledge but to generate new knowledge.

It is natural for linguists to focus on solving problems by improving the actual transcripts used. However, the framework offered here shows that the quality of the transcript may not be the only, or even the main, cause of problems. Further, where improved transcripts are needed, emulating the kinds of transcript used in linguistics may not be the best approach. As discussed in detail throughout this paper (especially Sections 2.4–5 and 5.2), a major finding traversing all branches of linguistics is that no transcript is universally valid: each must be tailored for its context. Legal contexts differ substantially from the contexts of traditional linguistics research. For example, in many legal contexts, even if the transcript is created by a linguist, it is used by a third party who interprets it under conditions not controlled by the linguist.

Transcription in legal contexts, then, requires accountable, evidence-based methods designed to ensure reliable interpretation in relation to their specific purposes and the specific conditions under which they will ultimately be used. Achieving this requires “end-to-end” research, that considers all the factors affecting the system as a whole. This poses new challenges for linguistics – and the high stakes of the criminal justice system means failure to meet them fully has serious consequences. Success in meeting the challenges, however, has value beyond legal contexts. Improved understanding of transcription as a general process promises benefits for the many other branches of linguistic science whose research depends on transcripts.

2. WHAT IS A TRANSCRIPT?

2.1. Transcription vs. Writing

A transcript is a representation of spoken language using the symbols of written language. It is important to distinguish transcription from writing, which itself is often taken to be a representation of spoken language. However, while this view is fostered by (and, arguably, needed for) primary literacy acquisition, it is not technically correct (Daniels and Bright, 1996). Writing and speaking are completely different ways of representing linguistic meaning (Ong, 1982). It is true that, to count as writing (as opposed to a picture, for example) a representation must have a systematic relationship to the sound system of the particular language it represents (DeFrancis, 1989). However, that relationship is indirect and partial – nothing like the direct representation of individual “sounds” with letters that many assume it to be on the basis of literacy education (Linell, 1988; Gillon, 2007).

A transcript, then, is unlike writing precisely in that it *does* aim to create a direct representation of the words (and sometimes the sounds, gestures or other elements) that were actually used by a speaker during a specific speech event – after that event has taken place. Interestingly, however, as discussed in detail below, no transcript can fully achieve this aim. A transcript gives a valuable way to recall and refer to spoken language, but can never substitute for the speech itself. A useful analogy (see Fraser and Loakes, 2020) is that a transcript is like a map. No map can ever give a full account of the territory it represents, and any map is valuable only to the extent it helps its end-users fulfil their needs. The same, this paper will argue, is true of transcripts.

2.2. Verbatim Reporting

While there are many forms of transcript, we can introduce some key concepts by starting with the simplest: the verbatim report. Verbatim reports aim to represent each speaker’s utterances, word by word, in ordinary spelling. They are now typically made from audio recordings. However, it is worthwhile to start by considering the traditional process: transcribing from live speech.

Writing down spoken language word by word seems simple in principle, but in practice it can be very hard. The most obvious difficulty is the speed at which spoken language is produced. No one can write quickly enough to capture all the words in real time – unless the speaker artificially slows down production, as in a schoolroom spelling exercise. At normal speaking rates, though a listener may recall the gist of what was said, the actual words are usually forgotten faster than they can be spelled out (Gurevich et al., 2010).

Transcription therefore requires an intermediate stage: creation of a temporary “record” of what was said, which can then be “written across” (the etymological meaning of “transcribe”) to create the “verbatim transcript”. The simplest way to make an intermediate record is by taking rough notes to use as an “aide memoire” (aid to memory). However, even with the aid of notes, it is hard to reconstruct the exact words the speakers used. Further, to the extent it can be done, there is no way to check for accuracy, except by comparing the memories – or notes – of other

participants. The resulting “transcript” has, at best, the character more of meeting minutes than of a verbatim record.

The need for accountable verbatim transcripts of official events led to development of special ways of capturing the intermediate record quickly and accurately: stenography (“narrow writing”) or shorthand. The skill of taking shorthand, and the techniques and procedures needed to transcribe shorthand into a text suitable for the readers who will eventually use it, were perfected over centuries, and professional stenographers have been in regular use in English courts, and other institutions, since the 1700s (Scharf, 1989). Since then, verbatim reporting has grown into the major world-wide industry our society relies upon today (e.g., intersteno.org). However, the increasing availability of practical audio recording techniques has seen reliance on stenographers gradually giving way to transcription from audio. Among other effects, this has highlighted some misconceptions about the nature of transcription.

2.3. Verbatim Transcripts From Audio

Those who have never tried transcribing from audio often assume it is easy, at least for a clear recording. After all, it solves the problem of speed faced by “live” transcribers. The audio captures a full record of exactly what was said, which can be paused and replayed at will, making transcription seem like a basic task, requiring little more than ability to spell.

The interesting thing is, however, that end-users often complain that the quality of transcription from audio is lower, not higher, than that of the apparently more difficult live transcription. The reason is that, on the assumption that “having the audio” makes transcription easy, managers tend to hire transcribers with lower qualifications than professional stenographers, and seek to increase output by farming work out to available transcribers, so that each transcribes short sections of multiple unrelated recordings.

The point is that, though the speed of speech may be the most obvious difficulty of transcription, it is not the only difficulty (Fraser, 2021a). So while the change to audio solves one problem, it creates others, especially by taking the speech out of its original context. The reasons are summarised in the next section; for extended discussion, see Fraser and Loakes (2020).

2.4. Transcription Is Not Transduction

The expectation that transcription should be easy reflects the everyday misconception that it is a mere transduction, in which words are mechanically copied from spoken to written form, and back again. This “transduction misconception” is incorrect, but nevertheless retains a powerful hold on common knowledge.

In this, it is similar to the widespread misconception that translating or interpreting from one language to another is a mechanical substitution of words in the source text with equivalent words of the target language. Actually, of course, translating and interpreting are complex skills, requiring many expert choices to be made in light of detailed understanding of the content and context of the material being translated (cf. Munday, 2016). That is why a translation is never “the” translation but always “a” translation – as demonstrated by the

fact that back-translation (translating a translation back into the original language) typically creates a text quite different from the original.

What is less commonly noted, though on reflection it is perfectly evident, is that reading a transcript aloud (a process that could reasonably be called “back transcription”) creates a speech event quite different from the original. This highlights the fact that a transcript, too, is never “the” transcript, but always “a” transcript. Speech is a massively complex signal, and it is impossible to represent it in its totality, even with specialised phonetic symbols (Heselwood, 2013). Transcribing speech into written text (like mapping a territory) requires many choices to be made regarding which elements to include, and how to represent them. Consider, for some simple examples: whether to include or omit false starts, self-corrections or hesitation markers; whether to represent colloquial or dialectal expressions with standard spelling or special symbols.

The effect is that any speech event can be represented in multiple ways, each with its own flavour. In fact, it is rare for two transcripts of the same material to be exactly the same. This gives linguists who teach transcription a handy way to detect cheating, as identical transcripts are likely to indicate that one has been copied from the other, despite student protests that they both independently “got it right”. Similar reasoning, in a far more serious context, is discussed by Coulthard et al. (2017) p. 116–120.

These and other considerations demonstrate that transcription from audio, far from being a simple transduction, is an especially complex form of symbolic representation, well named as “entextualisation”.

2.5. Entextualisation

The term “entextualisation” is relatively new (Urban, 1996; Park and Bucholtz, 2009), but the process has been researched for many decades (Ochs, 1979; Jefferson, 2004). One of the major findings is that producing verbatim transcripts requires context-sensitive interpretation by practitioners who are necessarily deeply embedded in specific social, cultural and political situations.

Much entextualisation research has focused on demonstrating that, despite this context-dependence, transcripts of official proceedings are often presented as “the” transcript – a manifestation of the transduction misconception that serves the interests of politically dominant elites, by treating the official transcript as objective, factual and neutral when really it reflects a particular point of view (Green et al., 1997; Roberts, 1997; Bucholtz, 2000).

This is important work – but the transduction misconception has other effects too. Erasing the role of the transcriber (Eugeni, 2020) diminishes respect for the many skills that professional transcribers bring to their task, meaning they may not receive the training and conditions they need to do an excellent job, as discussed above.

Another issue becomes particularly significant with transcription from audio. It is not only conscious choices that affect how words are represented. Context-sensitive interpretation, operating below the level of consciousness,

plays a far larger role in speech perception than most people realise. For a famous example, the same stretch of speech can be heard as “recognise speech” or “wreck a nice beach”, depending on the listener’s contextual understanding (see Fraser and Loakes, 2020). This is one of the factors that limited computer speech recognition in early decades. Development of practical systems had to await the technical ability to build contextual prediction into the programming (Pieraccini, 2012). Even now, automatic transcription, while valuable as a labour-saving measure, is typically only useful for relatively clear speech with well-separated turns (Loakes, 2022), and even then, accuracy requires careful editing by a human who understands the context and intended content (Love, 2020).

However, while the role of contextual information is by now well established in speech perception research, the ubiquity of the transduction misconception means that transcripts are often produced with inadequate control over the conditions that affect their quality. We have seen, for example, that working hour by hour on recordings from different trials simply does not allow a transcriber to build up sufficient contextual understanding. Similar issues are a major cause of the systemic problems that this paper seeks to address. Identifying and solving such problems requires recognising transcription as a skilled practice which takes place as part of a complex process involving context-sensitive interpretation at multiple levels, by practitioners in multiple roles. The next sections aim to contribute to this recognition, by suggesting a framework that sets out the main components of the complex process of transcription, and examining the factors that affect the quality of the resulting transcript.

3. A FRAMEWORK FOR UNDERSTANDING THE FACTORS THAT AFFECT THE QUALITY OF A TRANSCRIPT

The framework suggested here is based on the understanding, discussed above, that transcription requires three stages, which may be performed by different practitioners, or by one practitioner taking different roles:

- Stage 1: capturing an intermediate record;
- Stage 2: producing a transcript; and
- Stage 3: interpreting and using the transcript.

The reliability of a transcript is often attributed directly to the accuracy of the transcriber at Stage 2. However, it is important to pay explicit attention to all stages, each of which, as we will see, is subject to substantial misconceptions. In particular, each tends to be treated as transduction, when in fact all of them require context-sensitive, and often content-aware, interpretation.

Stages 1 and 2 require practitioners to “abstract” the information that seems relevant, in light of their understanding of the purpose of the transcript, from the overall context. This results in the “decontextualisation” of the transcript often emphasised in the entextualisation literature. One effect is that only information abstracted at earlier stages is available at later stages, making it easy for errors to propagate from one stage to

the next. In order to understand and use the decontextualised transcript, at Stage 3, the end-user has to “recontextualise” it, relying on knowledge, or assumptions, from various sources. Komter (2019) gives an especially clear account of these processes and their effects.

It is sometimes suggested that this reliance on context means transcription is necessarily “subjective” or even “biased”. However, these terms have multiple meanings, some with negative connotations which are not always appropriate for transcription. For example, “bias”, in its primary sense, suggests a conscious or unconscious intention to privilege interpretations that suit the practitioner’s interests. Bias in that sense can certainly affect any stage of transcription, with seriously undesirable consequences. That makes it essential to manage the transcription process so as to minimise opportunities for self-interest to be served. (The fictional account in Hannelore Cayre’s 2019 novella “The Godmother” gives an entertaining and not entirely implausible insight into the advantage an individual can take of a system with lax control.)

Managing bias has traditionally relied on security clearances and quality control. More recently, however, there has been a tendency to believe that it requires withholding contextual information from practitioners. This may be due to popularisation of the term “cognitive bias” for a range of psychological effects that do not necessarily involve self-interest (Kahneman, 2011). This usage has led some to believe that any context-awareness is necessarily biasing, and should therefore be eliminated. This is unfortunate. For reliable transcription, as for most other aspects of linguistic analysis, relevant, reliable contextual information is essential. Attempting to withhold all contextual information from practitioners can actually introduce biases of different kinds, which are even more difficult to manage effectively. The important thing, rather, is to ensure practitioners receive relevant and reliable contextual information, in a managed process, without exposure to potentially misleading information (cf. Dror et al., 2015).

Similar ambiguity surrounds use of the term “subjective”. Here the primary sense suggests personal preference influenced by an individual’s feelings or tastes – which is clearly not appropriate in scientific analysis. Avoiding subjectivity in this sense is often thought to require “objectivity”. The problem is that this term, too, has different interpretations. Often it is understood in the sense of requiring only context-independent measurement of observable physical features. However, by now it is well established that, even in the so-called “hard” sciences, observations and measurements are rarely fully “objective” in this strong sense (Hoffman, 2019; Ritchie, 2020). Almost all require human judgment (Kara, 2022). Trying to pretend they do not merely allows hidden biases to have uncontrolled and potentially damaging effects (D’Ignazio and Klein, 2020; Fry, 2021).

Striving for “objectivity” in that unrealistic – and outdated – sense, then, may be counterproductive for some sciences, especially for human sciences involving analysis of language. The important thing for scientific reliability in such fields is not to deny the role of human judgment, but to ensure that important judgments are made by a disinterested expert in relevant disciplines, who has full possession of relevant

reliable contextual information, carefully managed to preclude potentially misleading expectations, and can explain and justify their opinion in a transparent and accountable manner. To use the term “subjective” for the view of such an expert fails to distinguish it appropriately from a casual expression of personal preference. Perhaps some updated terminology is required in this area.

With these general remarks, we turn now to consideration of the factors that affect the overall enterprise of transcribing from audio, at each of its stages.

4. FACTORS AFFECTING THE CREATION AND USE OF TRANSCRIPTS

This section aims to set out some of the factors that affect the creation and use of transcripts of various kinds, with the focus on transcribing from an audio recording. The intention here is to present an overview for convenient reference, with examples and details in later sections. Of course, while it is useful to set the factors out separately, as this allows them to be considered methodically, they all interact extensively. The particular way they have been categorised here is influenced by the current focus on specific types of transcripts used in the legal process, and there are certainly other ways of conceptualising them (cf. Richardson et al., 2022). Indeed the present framework differs from, and supersedes, my own previous account (Fraser, 2014).

One key point that will be emphasised is that each factor involves expertise in a specialised field. Currently, few in linguistics have full expertise in all relevant fields, with a particular gulf between phonetics and other branches. Thus the discussion below does not claim to give definitive coverage of every factor, merely to indicate relevant considerations for each. Another key point is that all factors are heavily influenced by practitioners’ practical understanding of the purpose and context of their work at that stage – which can be influenced by knowledge or assumptions they may not be consciously aware of. In short, the output of each stage is never “the” output but only “an” output. However, though specialists in each factor are well aware of this fact, others have a strong tendency to over-simplify, with the transduction misconception being a particular problem through all stages.

4.1. Stage 1: Capturing the Audio Record

4.1.1. Audio Factors

Audio factors affect how the speech is abstracted from its context, and preserved for later listeners in an audio recording (with or without video). It is important to recognise that no audio is ever neutral. Like a photograph, a recording necessarily reflects the viewpoint of the one making it. So an essential overarching factor is the recording practitioner’s understanding of the purpose and context of the recording – which influences many decisions that affect the ultimate nature of the audio.

There are also numerous factors that affect the technical quality of the audio. These include the type of equipment being used, as well as the practitioner’s knowledge of how to use it, and ability to control how it is deployed. It is also important to

take account of any processing applied to the audio, whether at the time of recording, or later. For example, it is often assumed that “enhancing” indistinct audio makes it “clearer”, but this is not always true, and, again, the misconception can have negative consequences (Fraser, 2020a). For example, reducing background noise can have the undesirable effect of making listeners more, not less, likely to accept an inaccurate transcript (for a quick and compelling demonstration see Fraser, 2019).

4.1.2. Speech (and Speaker) Factors

Speech factors include the language, variety, register and style of the speech captured in the recording – all reflecting the speakers’ purpose, which, in almost all situations, is to make their meaning intelligible to intended or expected listeners. For “overt” (open) recordings, speakers may have awareness not just of listeners who are present at the time of the recording, but also of potential future listeners to the audio (cf. Haworth, 2013). In “covert” (secret) recordings speakers are typically aware only of the immediate listeners – though sophisticated criminals may consider possible hidden listeners, and attempt to disguise their meaning or identity.

An especially important factor is the location of the speech on the spectrum of formality. Informal conversation typically features overlapping and incomplete utterances, and is often highly elliptical, since listeners present at the time can rely for comprehension on implicit reference to aspects of the immediate context. However such references will be unavailable to those listening later to the decontextualised recording, potentially making the speech difficult to understand (video may help to some extent, assuming it is of good quality and designed to capture all relevant contextual information).

Since formal speech typically makes less reference to the immediate context, and is more likely to feature speakers taking separate turns, it may be intelligible even when technical quality is poor. Less formal conversation, however, may be heard inaccurately even with a good quality recording (Fraser and Loakes, 2020). A related factor is the pragmatic nature of the speech. For example, speech used for basic information exchange may be more readily represented in a verbatim transcript than nuanced social or emotional functions requiring subtle use of intonation and voice quality.

4.2. Stage 2: Producing the Transcript

4.2.1. Transcriber Factors

As we have seen, a recording is already an abstraction of the speech from its original context. Transcription involves further abstraction of the information needed to construct words and other linguistic entities from the recorded speech, and represent them in written form.

Perhaps the most obvious factor here is the practitioner’s level of training and testing in the technicalities of the specific style of transcript required. Equally important, though harder to test, is the practitioner’s personal aptitude for transcription. No transcript is ever “one and done”. All require significant concentration for repeated listening, with or without feedback from an evaluator (Section 4.2.3), and continual reviewing and updating of their work to reach a point of personal satisfaction

that it is of appropriate accuracy for the context. Another crucial factor, as always, is the transcriber's understanding of the purpose of the transcript, which affects many decisions about what aspects of the speech to include, and how to represent them.

4.2.2. Listener Factors

The "listener" here is not the listener to the original speech, but the listener to the recording. This is, of course, the same person as the transcriber, but in a different role. Indeed the listener role is arguably the most important role of all stages: after all, transcribers can only transcribe what they hear. Nevertheless it is one of the most overlooked roles of the entire transcription process, subject to many misconceptions.

One obvious factor is the listener's knowledge of the language, variety and register used by the speakers in the recording. Important as this is, however, it is only one factor – we cannot assume that anyone who knows a particular variety will automatically be good at transcribing any recording in that variety, especially if they have not been independently tested for aptitude under relevant conditions.

Another set of factors includes the listener's knowledge and expectations about the content and context of the recording, which, as outlined in Section 2.5 above, can have a large but typically unnoticed effect on perception, especially of audio with any degree of indistinctness. Again, however, while reliable contextual expectations can be helpful in understanding difficult audio, we cannot assume that those with reliable contextual knowledge will automatically create a reliable transcript – as this factor interacts strongly with aptitude and other factors.

A further important but little-recognised danger is that *unreliable* contextual expectations can be highly misleading, resulting in confident but inaccurate perception. Burrige (2017) gives a quick and accessible introduction to this concept, with entertaining examples showing just how easy it is for listeners to "hear" words that are not really there. Unfortunately, while examples like these are well known for their humour, their serious implications for transcription are not always fully recognised outside the specialised field of speech perception. This means that transcribers' contextual expectations are not always managed as diligently as they should be – a source of the problems discussed in Section 6.

4.2.3. Evaluator Factors

As mentioned above, a certain amount of personal evaluation is undertaken as part of the transcriber role. Some transcription situations also require external evaluation of the transcript, e.g., via a test used for accreditation or quality control. In such cases, there are additional factors to consider. One, clearly, is the evaluator's independence, understanding of their role, and knowledge of the factors that might influence their judgement.

Appropriate decisions about details of the test are also crucial. For example, it matters what the transcript is evaluated against – e.g., a known correct transcript, the evaluator's memory of what was said, or the audio itself. Particularly difficult issues arise in the last situation, since the very act of viewing the transcript in order to check it can affect the listener's interpretation of the audio (Section 6.1.1). Unfortunately, however, while the role of

such decisions is well understood in language testing (e.g., Knoch and Macqueen, 2020), transcript evaluation has not yet developed a sophisticated methodology.

4.3. Stage 3: Using the Transcript

4.3.1. End-User Factors

Another often-overlooked consideration is how the eventual transcript is actually used in practice by its end-user (the linguist, lawyer, jury, etc., who ultimately interprets its content). After all, even the best transcript can be used wrongly or inappropriately (just as an excellent map can fail if the end-user does not understand its capabilities and limitations – see Section 2.1).

The first factor to consider, as always, is the end-user's intention and purpose in using the transcript – which may or may not be the same as the intention and purpose of practitioners at other stages. Another is the end-user's understanding of the nature of transcription in general. Are they simply picking up "a" transcript and treating it as "the" transcript? Or are they considering appropriately whether this particular transcript is suitable for their purpose? If the latter, do they have sufficient knowledge of the transcript's provenance to be able to assess its suitability, and take account of its (inevitable) limitations? Finally, the end-user's ability to interpret any specific transcription conventions is important.

4.3.2. Overall System-Design Factors

Considering end-user factors raises the need to consider the transcription process as a whole, by evaluating the factors that affect each stage, and assessing the extent to which the overall system is working as intended. Ideally this would be done as part of the design and management of a system created in pursuit of a unified overall purpose, with appropriate consultation of those with expertise relevant to each stage. Alternatively, it could be done "post hoc", by retrospectively reviewing the factors that have contributed to the quality of the transcript and the end-user's ability to use it appropriately. Either way, it should be undertaken with full understanding of the expertise that is required of practitioners at each stage, and all the factors that contribute to the output.

However it can happen that neither of these kinds of system evaluation are undertaken effectively – or at all. Section 6 considers two such situations: transcripts of police interviews and forensic audio, and their propensity to induce errors with far-reaching negative implications for our criminal justice system. First, however, we consider two situations where the transcription process is (with important exceptions) designed, evaluated and used well: court transcripts and research transcripts. This will help in determining the key factors that contribute to successful creation and use of transcripts.

5. USING THE FRAMEWORK: TWO (GENERALLY) SUCCESSFUL EXAMPLES

This section demonstrates use of the framework by looking at two kinds of transcripts that serve very different purposes: transcripts of court proceedings, and transcripts used in linguistics research. In each case, the transcripts are generally successful in serving

their purpose – though, as we will see, both are subject to serious failings if particular factors are not managed appropriately. Discussion will demonstrate that success arises not from any single factor, but from pursuit of the transcript's overall purpose in light of well-informed, context-aware management of all relevant factors, along with careful, ongoing system evaluation.

5.1. Transcripts of Court Proceedings

The overall purpose of court transcripts is to create an official record of trial proceedings that can be used by anyone, and is trusted by all. Here we briefly consider the factors that affect the outcome, focusing first on the traditionally monolingual situation of Australia and the UK.

Most of the key speakers in a trial use relatively standard English, though individual witnesses may have a range of different dialects (witnesses who speak languages other than English are provided with an interpreter – at least in principle, if not always in practice: e.g., Cooke, 2009). Most speakers also use relatively formal language, monitored by the judge to ensure that everyone talks in turn, and all speak up clearly “for the tape”. Much of the speech involves basic information exchange – with departures from this usually evident from subsequent turns.

The audio quality is typically fair. Together these factors mean the recording is mostly easily intelligible by transcribers familiar with the courtroom genre, though listeners may have difficulty in making out unfamiliar names or technical terms.

Court transcribers are accredited to ensure they have the necessary skills for accurate verbatim transcription, and undergo security clearance to ensure their independence in relation to trial outcomes. They are also highly trained in the use of specific conventions appropriate to court transcripts, including how to “tidy up” the representation of spoken language (e.g., by eliminating hesitation markers or false starts) to make it easier for end-users to read, and to give a respectful impression of court-room discourse (cf. Voutilainen, 2018).

The transcriber in the role of listener typically knows the language, variety and register of the court (though not necessarily those of all witnesses, as noted below), and is provided with names and technical terms, as well as general contextual information, to assist in perception of unpredictable content. Evaluation of individual transcripts is undertaken by the lawyers and judges who took part in the trial – in light of their memory of what took place, and their understanding of what information court transcripts should capture. The end users are readers who understand the transcription conventions and the courtroom context. As mentioned earlier, the overall system has been designed over centuries with ongoing evaluation and development aimed at ensuring that court transcripts meet the needs of society, or at least of its dominant sectors (cf. Section 2.5).

Not surprisingly, given all these circumstances, courtroom transcripts are, in general, well suited to their purpose, and mostly of high quality – at least in the monolingual scenario for which the factors have been optimised. The fact that substantial problems have been demonstrated in representing the speech of witnesses with non-standard dialects (Walsh, 1995; Jones et al., 2019) shows that court transcription processes, despite their long

history, have been designed without full understanding of all relevant factors.

What is interesting to note now is that their general suitability for their own purpose does not imply that court transcripts are universally suitable for every purpose. In particular, they have substantial limitations when used as the basis of linguistic research on courtroom interaction, as discussed next.

5.2. Transcripts for Linguistic Research

Transcripts are used in many branches of linguistic research (some mentioned in Section 1 above). One that is of relevance here, and will enable exemplification of some general issues, is research on spoken interaction in court – aiming, for example, to demonstrate and theorise practices that create systematic disadvantage for certain categories of defendants (e.g., Eades, 2010; Mariottini, 2017).

The interesting thing is that court transcripts are generally not useful for this kind of research – precisely because they are not, in fact, strictly “verbatim” in the sense of representing each word as it was spoken (Eades, 1996). The “tidying up” undertaken by court reporters, though useful to intended end-users, can alter the very detail needed for the research. For this reason, researchers often choose to make their own transcripts – which of course are affected by their own set of factors.

Some factors are the same as for court transcripts. Research on courtroom interaction typically uses the courtroom recording, and the transcriber in the role of listener almost always knows the content with considerable certainty – as is true for almost all linguistic research.

Where the two differ sharply, however, is in the overall purpose of the transcript. Research transcripts aim, not to preserve the informational content of the speech for use by a generalised third party, but to represent and operationalise features of the spoken language for use by the transcriber (or close associates) in exploring whatever theoretical issues are under consideration. Thus while court transcripts are an end in themselves, linguistic transcripts are a means to an end: after peer review and publication, the transcripts themselves are rarely referred to again, unless to critique the research.

The transcriber is trained to focus on aspects of spoken language relevant to the research, and to annotate them via special formatting and technical symbols whose meaning and use must be learned via advanced education. Very importantly, however, these technicalities are an addition to, not a substitute for, reliable representation of the verbatim content. While technical symbols may impress outsiders, they can mask errors that reduce the overall reliability of the transcript. Also importantly, use of technical symbols does not imply the transcript is “objective” in the sense of being unbiased or neutral. It has long been known that research transcripts can display self-interested bias (Wald, 1995). For this reason, transcripts used in high-stakes research are usually subject to external evaluation, typically via inter-rater reliability checks, which compare transcripts from several transcribers, each with relevant expertise and knowledge of the overall purpose of the research – but “blinded” as to context that might engender bias.

5.3. Discussion

Both court and research transcripts are highly successful in their own domains – though not infallible, as we have seen. Indeed, the success of each comes precisely from its recognition of the potential for error, which motivates management of known risk factors, and commitment to ongoing independent evaluation and improvement of the system.

However, while these two types of transcript are successful in their own domains, they are very different – and not interchangeable. We have seen that court transcripts are generally not useful for linguistic research. Less obviously, perhaps, research transcripts are not useful as court transcripts. Importantly, this is not only because court transcribers and end-users lack the skills needed to produce and understand technical linguistic representations. Linguistic transcripts, like any others, require choices to be made, in light of context-aware understanding of their overall purpose, about what detail to include, and how to represent it. That is why linguists' transcripts can rarely be transferred from one research project to another (Jenks, 2013) – further reinforcement of the key insight, discussed above, that no transcript is a neutral representation.

This is important to emphasise here in light of the persistent misconception that certain kinds of technical transcripts can somehow capture the “objective truth” of what was said via “bottom-up” analysis. Such claims are sometimes made, for example, in relation to conversation analysis (CA). It may well be true that CA practitioners pursue data-focused analysis more diligently than some more “theory-driven” branches of linguistics. But this does not mean that CA transcripts are “neutral”, or “objective” in the strong and outdated sense discussed in Section 3 – as CA experts themselves are at pains to acknowledge (Edwards, 2008; Hepburn and Bolden, 2012).

Even stronger claims of “objectivity” in the outdated sense are made for phonetic transcription. Again, however, experts are clear that such claims are overblown (Heselwood, 2013; Himmelmann, 2018). Indeed one of the best established findings of speech perception research is that “bottom up” word recognition is impossible. That is why, for example, expert phoneticians acknowledge that they have limited ability to transcribe languages they do not know, or to “read” spectrograms with unknown content (see Fraser, 2022 for extended discussion).

Of course, this is not to suggest that either of these kinds of transcription are “subjective” in the soft sense of reflecting mere personal preference. Nor does it suggest that not being “objective” in the outdated sense diminishes the value of CA or phonetic transcripts. To the contrary – both are highly valuable in the contexts for which they are developed. What is essential, however, is to acknowledge that valid use of their specialised symbols depends crucially on valid understanding, both of the context and content of the audio, and of the purpose of the transcript, being shared by both creator and interpreter of the transcript.

What makes a transcript reliable and useful, then, is expert judgment, exercised across all three stages, in a system designed to manage the complex intertwined factors that affect the suitability of the final product to the end-user's needs. It is this type of management that makes both linguistic and court transcripts successful – and it is in being the product of this kind

of management that these two types of transcripts are similar, despite their many differences of style, content, layout, etc.

6. USING THE FRAMEWORK: TWO PROBLEMATIC EXAMPLES

With the insights of Section 5 in mind, it is now time to consider our two examples of transcripts being used in more problematic ways. Both forensic audio and police interviews start life as part of a criminal investigation, during which transcripts are used, if at all, in relatively unproblematic ways. Both, however, sometimes go on to serve as evidence in court, where transcripts can be used in ways that have been shown to create major problems for justice. This section aims to describe these problems, identify the factors that cause them, in light of the insights developed above, and discuss potential solutions.

The key observation will be that, while there has been an understandable tendency to focus on the transcriber as the main source of the problems, actually transcriber factors are only one part of the problem, and not necessarily the most important. So while expertise in linguistic science is essential to developing a better system for transcribing forensic audio, the expertise needed is not simply the ability to create technical linguistic transcripts. Rather expertise is needed to develop and manage an overall system that emulates, at a deep level, the practices that create successful transcripts – paying attention to all the factors, not just the superficial factor of being able to use technical symbols and terminology (Fraser, 2020c).

6.1. Transcripts of Indistinct Forensic Audio

Forensic audio is speech that has been captured, typically in a covert (secret) recording obtained as part of a criminal investigation, and is later used as evidence in a trial. Such recordings provide powerful evidence, allowing the court to hear speakers making admissions they would not make openly. One problem, however, is that the audio is often extremely indistinct, to the extent of being unintelligible without the assistance of a transcript.

Transcripts used to give this assistance are typically provided by police investigating the case, who, in court, are given the status of “ad hoc expert” on the grounds that they have listened to the audio many times. This is often found alarming by linguists, who suggest it would be better to have the transcripts produced by real experts. Surprisingly, however, insisting on expert transcripts, though surely an improvement, is not a fool-proof solution (Fraser, 2020b, 2021b). To gain an impression of the reasons, and to consider directions to look for better solutions, it is worth reviewing the factors that cause problems with police transcripts.

6.1.1. Factors Affecting the Reliability of Police Transcripts of Forensic Audio

The combination of very poor technical quality, and unmonitored, highly contextualised conversation means many covert recordings are essentially unintelligible to general listeners. The purpose of the transcript is to assist the court in perceiving the content, and thus in better understanding the context (i.e. the crime, and who is responsible for it).

Ad hoc experts have no training in transcription, and are not required to demonstrate skill. The reason they are asked to provide transcripts has to do with their role, not as transcriber, but as listener: they can often make out more of the content of indistinct audio related to their cases than other listeners can. Though the law attributes this ability to their having listened many times, the real reason is their access to contextual information – and it is important to acknowledge that reliable contextual information can sometimes help police understand specific utterances. As discussed in Section 4.2.2, however, mere access to contextual information cannot guarantee a reliable transcript. A particularly serious limitation on police transcripts is that not all contextual information available to investigators is reliable (that is why we need the trial). The powerful effect of contextual expectations on perception means that unreliable contextual information can easily mislead perception, without conscious awareness. For these reasons, police transcripts are rarely fully accurate, and often egregiously wrong (French and Fraser, 2018).

The end-user is the jury, who are instructed by the judge to listen carefully to the audio and form their own opinion as to its content, using the transcript only as assistance. Unfortunately, however, this is an unrealistic instruction. It is well known that an inaccurate transcript can easily “assist” listeners to hear words that are not there (Section 4.2.2). Indeed, the law is aware that police transcripts might be wrong, and a transcript is not provided as assistance to the jury until it has been evaluated. The problem is that the evaluation is carried out by lawyers checking the transcript against the indistinct audio, without realising that this very process inevitably subjects their own perception to the influence of a potentially misleading transcript (Fraser, 2018; Fraser and Kinoshita, 2021).

Finally, the overall system has been designed by judges, on the basis of their experience with court transcripts, with insufficient understanding of the factors that influence understanding of indistinct forensic audio. No system evaluation is undertaken. The whole process is driven, not by scientific values, but by legal precedent (Fraser, 2021b).

6.1.2. Discussion

Unsurprisingly, this process gives rise to serious problems, and numerous instances of injustice have emerged (for a quick introduction with an interesting connection to Section 6.2, see Fraser, 2013). However setting out the factors methodically has shown that the main cause of these problems is not the fact that transcripts are provided by investigators (though this is far from ideal). The problems are created by the system as a whole, with the most important factor being the fact that transcripts of indistinct forensic audio are evaluated by lawyers involved in the trial. Even transcripts provided by experts are evaluated by lawyers and judges, creating substantial problems (Fraser, 2021b). So the first step towards improvement must be to change the legal procedures that give so much credence to inexpert and unaccountable evaluation of transcripts (Fraser, 2020c).

The next step is to introduce processes for providing courts with reliable transcripts. Many have assumed that this can be achieved by individual experts evaluating police transcripts -

as I did myself until casework experience led me to argue this it is not suitable, for a range of reasons (Fraser, 2020b). These reasons have recently been amplified by a ground-breaking study (Love and Wright, 2021) in which eight different (expert) transcribers of indistinct audio created eight transcripts that differ in substantial ways. The point is that the experts were operating under uncertainty regarding the true content of the audio. This of course is the standard situation with forensic audio – but very different from any kind of linguistic research (Section 5.2). Further, while acoustic analysis might confirm some parts as more or less likely to be right, the true content is unlikely to be established purely by “bottom up” analysis (Section 5.3). These differences clearly indicate a need for specialised system design.

Producing a reliable transcript of indistinct audio of unknown content needs methods beyond standard linguistic or acoustic analysis. To date, however, very little research has been directed explicitly towards developing such methods (see Fraser, 2022). New projects are needed to design an evidence-based process that can ensure all forensic audio used in court is provided with a reliable transcript (or certified as incapable of reliable transcription). Such projects need to take an end-to-end approach, to ensure the transcripts are suitable for the purpose of assisting a jury to understand the content under courtroom conditions (recognising there can be a major difference between the information an expert puts into a transcript, and the information end-users take from it).

We cannot leave this section without mentioning that indistinct covert recordings frequently feature languages other than English, which require not only reliable transcription, but also reliable translation. Unfortunately both of these tasks are carried out according to procedures developed with poor understanding of relevant aspects of linguistic science (Fraser, 2021b). Even more unfortunately, valuable efforts of experts to document the resulting problems (Capus and Griebel, 2021; Gilbert and Heydon, 2021) and suggest viable solutions (González et al., 2012; NAJIT, 2019) are so far having limited impact on general practice.

6.2. Transcripts of Police Interviews With Suspects

We turn now to our second problematic example: transcripts of police interviews with suspects. Traditionally, these were created on the basis of an intermediate record made by officers taking notes about what the suspect said (cf. Section 2.2 above). This famously gave opportunities for “verballing” – police falsely claiming that suspects had made “verbal admissions” during the interview (Eades, 2010; Grant, 2022). In both Australia and the UK, Royal Commissions in the 1980s and 1990s sought to curtail opportunities for such “fabricated confessions”, by instituting requirements that all police interviews with suspects should be audio/video recorded (Baldwin, 1985; Dixon, 2008). This is now gradually being extended to an expectation that police will use body-worn recording devices while interviewing witnesses or engaged in other duties (Roberts and Ormerod, 2021).

Electronically recorded interviews have many benefits. One disadvantage, however, is that recordings are not convenient to access or refer to. This makes it necessary to provide a transcript of each interview. Upon institution of compulsory recording, the large workforce needed for transcription was mobilised hastily and under severe cost constraints, often co-opting practitioners whose primary skills and responsibilities lay elsewhere. Unfortunately it was not till decades later that it was discovered that their transcripts sometimes contained egregious but undetected errors, with potential to affect justice (Haworth, 2018; Komter, 2019; Richardson et al., 2022).

Again, before considering solutions to this problem, it is useful to review the factors methodically, so as to ensure its key causes are identified properly.

6.2.1. Factors Affecting the Reliability of Police Interviews With Suspects

The audio quality of recorded police interviews is usually fair, and the style of speech is usually relatively formal and relatively well monitored. This means that the audio is usually reasonably intelligible – though typically well below the standard of recordings of court proceedings, making the task of interview transcribers harder than that of court transcribers. The audio quality of body-worn recordings can be particularly poor.

Despite the harder task they face, interview transcribers are rarely as well-qualified, nor as well-resourced, as court reporters. The fact that they are typically employed by police departments, or by agencies that undertake extensive police work, means they usually have contextual understanding of police and legal processes in general, and sometimes of specific cases. Nevertheless, various kinds of error are common, as well documented by Haworth (2018) and Komter (2019) – confirming that difficulties in understanding recorded speech are not limited to poor quality audio (Section 4.1.2).

Evaluation of interview transcripts is effectively non-existent. In principle, it is intended to be undertaken by lawyers, with the defence considered especially responsible for reviewing the transcript, as shown by the following advice for defence lawyers:

It is important to watch the [video] or listen to audio tapes of records of interview. It will not only help you work out whether the transcript is accurate, but it may also indicate important aspects of the questioning and your client's manner and condition at the time of questioning which may be relevant in your case (for example, being intoxicated or not in a fit mental state) (NSW Young Lawyers Criminal Law Committee, 2004: 172).

Evaluation of transcripts by lawyers is not ideal, since they have neither the expertise nor the independence to undertake the task rigorously, making it unlikely that they would detect all relevant errors. Worse still, even this less-than-ideal evaluation is often skipped. Time pressures mean the advice below is not always followed – making it common for the transcript to be used as the definitive account of the interview, with the audio never being accessed at all, let alone used for careful evaluation of the transcript.

Copies of your client's [recording] will not usually be included in the prosecution brief. You will generally be served only with a transcript of what was said in the [interview]. You should get a copy of your client's [recording] (NSW Young Lawyers Criminal Law Committee, 2004 p.284).

The end-user is the most complex factor in this situation. Typically, multiple parties use the transcript (cf. Haworth, 2013) – each with different needs. First, the police themselves may use it to aid their memory of what happened in the interview (though they may prefer their own notes). Then prosecution and defence solicitors use it, in preparing their cases, as a record of the information obtained during the interview. Next, if the interview is used as evidence in court, barristers quote from the transcript, using their own intonation and speaking style (Haworth, 2018). The final, and arguably most important, end-user, is the jury, who use the content of the interview, in combination with other evidence, to reach a verdict of guilty or not guilty. As is clear from the above account, however, they may understand the content only through a barrister's "back-transcription" (Section 2.4). Unlike the situation with forensic audio, there is no expectation or requirement that interview audio be played in court.

System design and evaluation are close to non-existent. Developed in haste, and with no input from relevant experts, the whole process was subject to little scrutiny until researchers like Haworth and Komter exposed some of its serious weaknesses:

[I]n stark contrast to the strict principles of preservation applied to physical evidence, interview data go through significant transformation between their creation in the interview room and their presentation in the courtroom, especially through changes in format between written and spoken text (Haworth, 2018: 428).

6.2.2. Discussion

As with forensic audio, it is common for the failings of interview transcripts to be blamed on the transcriber. Again, however, it is clear from the above analysis that the problems lie in the system as a whole, which is designed and managed with insufficient attention to crucial factors. This means that the problems cannot be solved purely by seeking ways to ensure more reliable transcripts (though this is certainly an important part of the solution, as discussed shortly). After all, even an excellent transcript risks giving a misleading impression of the audio if it is read out by a barrister, selectively using intonation, pausing, etc., designed to persuade a jury to accept a particular version of what happened in the interview. Preventing this would seem to require working with the judiciary to reform practices for presenting interviews as evidence in courts – by demonstrating how essential it is for the court to listen to the actual audio.

Further, as discussed above, interview transcripts are not always excellent. It is really essential to ensure they are always of high quality. The question is how to achieve this. One common suggestion is to train interview transcribers to include more detail in their transcripts, perhaps creating a simplified version of the style of transcript used in branches of linguistics like conversation analysis (CA). However this suggestion raises several issues.

First, the value of a CA-style transcript is limited by the accuracy of the verbatim representation on which it is based (Section 5.3). If verbatim transcripts contain errors, adding technical detail will not help – and may actually mask deficiencies by making it even more difficult for listeners checking the transcript against the audio to notice errors (Section 4.3). The priority then, might be to ensure that interview transcribers produce reliable verbatim transcripts – not by insisting busy lawyers check the transcript against the audio, but by training, resourcing and managing interview transcribers in ways commensurate with courtroom transcribers (Section 5.1).

Second, learning even simplified CA transcription is difficult, especially for transcribers with no background in linguistics. While they may be taught some technicalities, they may retain misconceptions about language and speech that undermine their ability to use the teaching effectively (at least, this is a common outcome when training in phonetics is provided to assist English pronunciation teachers, see Burri et al., 2017).

Third, the detail in a CA transcript necessarily reflects the transcriber's understanding of its context and purpose (Section 5.3). This is not a problem for research transcripts, where end-users share the same context and purpose as transcribers. With interviews, however, end-users (especially lawyers on opposing sides) need to form their own independent interpretation of the interview in light of their own purposes, with minimal influence from the interpretations of others.

Finally, and most importantly, no transcript can represent all the information in the audio, as discussed at length above. Using any transcript, even one with detailed and accurate annotation, without reference to the audio, inevitably causes end-users to miss or misinterpret aspects of the content – as has now been powerfully demonstrated, specifically in relation to police interviews, by Deamer et al. (in press). In a worst-case scenario, an annotated transcript could even serve, intentionally or not, to manipulate end-users' understanding of what was said in the interview, especially when speech is nuanced, emotional or otherwise open to varying interpretation.

For all these reasons and more, it is really essential for end-users of interview transcripts to listen to the recording personally. Unfortunately, as we have seen, this rarely happens. While one reason is time-poverty, another is the transduction misconception. Lawyers on both sides simply accept that the transcript is essentially equivalent to the audio:

[contamination of interview data] appears to stem from a lack of recognition that changes in the format of linguistic data involve transformation of the data themselves. A first step in improving current practice, then, is to increase awareness of that simple fact (Haworth, 2018: 445).

To persuade busy lawyers to listen to the audio, then, one approach might be to institute education, especially for those on the defence side, in which linguists can explain the falsity of the transduction misconception, and demonstrate how listening to the audio can reveal information that might help win a case – hopefully thus motivating solicitors to request video recordings at the start of each case (or, better still, to get them routinely without need for a request).

To make the listening more efficient, it may be worth noting that substantial proportions of police interviews are taken up with routine information-exchange, which can be understood relatively well from a standard verbatim transcript (Section 4.1.2). One suggestion worth exploring, then, might be to ask transcribers to draw the attention of lawyer end-users to parts that most need to be listened to, simply via marginal notes indicating sections of the transcript where the language diverges, in any way, from straightforward information-giving. This takes less skill, and less interpretation, than a detailed CA transcript, but could help busy solicitors to use their listening time for the most salient parts of the interview. Of course it would be necessary to test this suggestion via ecologically valid, end-to-end research, involving linguists, transcribers and lawyers, to discover whether it works well in practice. If it does, ongoing training and management would be needed to maintain appropriate standards (cf. Richardson et al., 2022).

Finally, as before, it is impossible to leave this section without mentioning the topic of interviews that involve languages other than English. Linguists are already well aware of poor practice in communication during interviews between police and less proficient speakers of English (e.g., Eades, 2018; Bowen, 2021), and are undertaking valuable research to bring improvement (e.g., Hale et al., 2019). It is certain there must also be major issues in relation to how transcripts of interpreted interviews are produced and used (cf. NAJIT, 2019). However, to my knowledge little has yet been done even to document these issues (though see Gibbons, 1995), let alone to solve them. Of course, interviews requiring use of Deaf sign language raise their own issues.

7. CONCLUSION

This systematic review started by discussing the nature of transcription, and setting out a framework for understanding the factors that affect a transcript's reliability and suitability for purpose. It then demonstrated how the framework can explain the successful use of two types of transcript that superficially appear to share few characteristics in common, namely court transcripts and transcripts used in linguistic research. This demonstration emphasised that a transcript is not the product of an individual transcriber working in isolation, but of a range of roles and factors that interact in complex ways. Ensuring the reliability and usability of a transcript requires managing all of these roles and factors effectively, with good understanding of how the transcript will ultimately be interpreted by the end-user. It is successful management at this level that ensures the success of court transcripts and linguistic transcripts for their disparate purposes.

The review then turned to two fields in which use of transcripts has been shown to be highly problematic, namely forensic audio and police interviews used as evidence in court. Emphasising that solving the problems with these transcripts requires careful identification of exactly what causes the problems, it then subjected each to analysis of the factors indicated by the framework. This showed that in neither case

can the problems be addressed effectively simply by bringing the transcripts more into line with those used in linguistics research. Developing effective solutions requires considering high-level system-design factors, especially the transcript's overall purpose, and the conditions under which end-users interpret it.

This suggests a need for two strands of research, one directed towards improving provision of transcripts in a range of legal contexts, and another directed towards improving legal procedures, to ensure that good transcripts, once available, are used well. An excellent model for this kind of double-stranded research-based engagement between linguists and judges is provided by development of the Australian *Recommended National Standards for Working with Interpreters in Courts and Tribunals* (JCCD, 2022) – already used as inspiration in seeking improvement for transcripts of forensic audio (Fraser, 2020c).

It is hoped that the analysis offered in this systematic review will contribute to improving transcription in all legal contexts. A further hope, however, is that the “framework for deciding how to create and evaluate transcripts for forensic and other purposes” offered here, suitably amended via interdisciplinary discussion, might also be applied more broadly, helping to consolidate transcription as a dedicated field of study within linguistic science. After all, transcripts form the

foundation of a large proportion of research in many branches of linguistics.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

This paper, even more than most academic research, stands on the shoulders of previous scholars, namely those many linguists who for decades have sought to overcome societal misconceptions about the capabilities and limitations of transcripts as a representation of speech. I particularly acknowledge the work of Kate Haworth, whose publications on police interviews helped me see transcription in legal contexts as a general issue, beyond forensic audio, which could benefit from a framework like the one offered here. I also acknowledge valuable ongoing discussion with generous colleagues, especially Debbie Loakes; and careful, constructive comments from two reviewers. Remaining shortcomings of the framework, which I hope will be improved via collegial debate, are of course my own responsibility.

REFERENCES

- Baldwin, J. (1985). The police and tape recorders. *Crim. Law Rev.* 695–704.
- Bowen, A. (2021). Intercultural translation of vague legal language: the right to silence in the Northern Territory of Australia. *Target. Int. J. Transl. Stud.* 33, 308–340. doi: 10.1075/target.19181.bow
- Bucholtz, M. (2000). The politics of transcription. *J. Pragmat.* 32, 1439–1456. doi: 10.1016/S0378-2166(99)00094-6
- Bucholtz, M. (2009). Captured on tape: professional hearing and competing entextualizations in the criminal justice system. *Text Talk Interdiscip. J. Lang. Discourse Commun. Stud.* 29, 503–523. doi: 10.1515/TEXT.2009.027
- Burri, M., Baker, A., and Chen, H. (2017). “I feel like having a nervous breakdown”: pre-service and in-service teachers’ developing beliefs and knowledge about pronunciation instruction. *J. Second Lang. Pronunc.* 3, 109–135. doi: 10.1075/jslp.3.1.05bur
- Burridge, K. (2017). The dark side of mondegreens: how a simple mishearing can lead to wrongful conviction. *The Conversation*. Available online at: <http://theconversation.com/the-dark-side-of-mondegreens-how-a-simple-mishearing-can-lead-to-wrongful-conviction-78466> (accessed June 26, 2022).
- Capus, N., and Griebel, C. (2021). The (in-)visibility of interpreters in legal wiretapping. *Int. J. Lang. Law* 10, 73–98. doi: 10.14762/111.2021.73
- Cooke, M. (2009). Anglo/Aboriginal communication in the criminal justice process: a collective responsibility. *J. Judic. Adm.* 19, 26–35
- Coulthard, M., Johnson, A., and Wright, D. (2017). *An Introduction to Forensic Linguistics: Language in Evidence, 2nd Edn.* London/New York, NY: Routledge.
- Coulthard, M., May, A., and Sousa-Silva, R. (eds.). (2020). *The Routledge Handbook of Forensic Linguistics*, 2nd Edn. London/New York, NY: Routledge. doi: 10.4324/9780429030581
- Daniels, P., and Bright, W. (1996). *The World's Writing Systems*. Oxford: Oxford University Press.
- Deamer, F., Richardson, E., Basu, N., and Haworth, K. (in press). Exploring variability in interview interpretations. *Language and Law/Linguagem e Direito*.
- DeFrancis, J. (1989). *Visible Speech: The Diverse Oneness of Writing Systems*. Honolulu, HI: University of Hawaii Press.
- D'Ignazio, C., and Klein, L. (2020). *Data Feminism*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11805.001.0001
- Dixon, D. (2008). *Videotaping Police Interrogation*. University of New South Wales Faculty of Law Research Series, p. 28.
- Dror, I., Thompson, W., Meissner, C., Kornfield, I., Krane, D., Saks, M., et al. (2015). Context management toolbox: a linear sequential unmasking (LSU) approach for minimizing cognitive bias in forensic decision making. *J. Forensic Sci.* 60, 1111–1112. doi: 10.1111/1556-4029.12805
- Eades, D. (1996). “Verbatim courtroom transcripts and discourse analysis,” in *Recent Developments in Forensic Linguistics*, ed H. Kniffka. Bern: Peter Lang. p. 241–254.
- Eades, D. (2010). *Sociolinguistics and the Legal Process*. Bristol: Multilingual Matters. doi: 10.21832/9781847692559
- Eades, D. (2018). Communicating the right to silence to Aboriginal suspects: lessons from Western Australia v Gibson. *J. Judic. Adm.* 28, 4–21
- Edwards, J. (2008). “The transcription of discourse,” in *The Handbook of Discourse Analysis*, eds D. Schiffrin, D. Tannen, and H. Hamilton (Oxford: Blackwell Publishing Ltd), p. 321–348.
- Eugeni, C. (2020). The reporter's invisibility. *Tiro J. Prof. Report. Trans.* 2.
- Fraser, H. (2013). Covert recordings as evidence in court: the return of police ‘verballing’? *The Conversation*. Available online at: <https://theconversation.com/covert-recordings-as-evidence-in-court-the-return-of-police-verballing-14072> (accessed June 26, 2022).
- Fraser, H. (2014). Transcription of indistinct forensic recordings: problems and solutions from the perspective of phonetic science. *Lang. Law Linguagem e Direito* 1, 5–21.
- Fraser, H. (2018). Forensic transcription: How confident false beliefs about language and speech threaten the right to a fair trial in Australia. *Aust. J. Linguist.* 38, 586–606. doi: 10.1080/07268602.2018.1510760
- Fraser, H. (2019). Don't believe your ears: “Enhancing” forensic audio can mislead juries in criminal trials. *The Conversation*. Available online at: <https://theconversation.com/dont-believe-your-ears-enhancing-forensic-audio-can-mislead-juries-in-criminal-trials-113844> (accessed June 26, 2022).
- Fraser, H. (2020a). Enhancing forensic audio: what works, what doesn't, and why. *Griffith J. Law Hum. Dign.* 8, 85–102.

- Fraser, H. (2020b). "Forensic transcription: the case for transcription as a dedicated area of linguistic science," in *The Routledge Handbook of Forensic Linguistics*, eds M. Coulthard, A. May, and R. Sousa-Silva (London/New York, NY: Routledge), 416–431. doi: 10.4324/9780429030581-33
- Fraser, H. (2020c). Introducing the research hub for language in forensic evidence. *Judic. Officers Bull.* 32, 117–118.
- Fraser, H. (2021a). How misconceptions about transcription affect the criminal justice system. *Tiro J. Profess. Report. Transc.* 3.
- Fraser, H. (2021b). The development of legal procedures for using a transcript to assist the jury in understanding indistinct covert recordings used as evidence in Australian criminal trials: a history in three key cases. *Lang. Law Linguagem e Direito* 8, 59–75. doi: 10.21747/21833745/lanlaw/8_1a4
- Fraser, H. (2022). "Forensic transcription: legal and scientific perspectives," in *Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications*, eds C. Bernardasci, D. Dipino, D. Garassino, E. Pellegrino, S. Negrinelli, and S. Schmid (Milano: Officinaventuno), 19–32.
- Fraser, H., and Kinoshita, Y. (2021). Injustice arising from the unnoticed power of priming: how lawyers and even judges can be misled by unreliable transcripts of indistinct forensic audio. *Crim. Law J.* 45, 142–152.
- Fraser, H., and Loakes, D. (2020). Acoustic injustice: the experience of listening to indistinct covert recordings presented as evidence in court. *Law Text Cult.* 24, 405–429.
- French, P., and Fraser, H. (2018). Why "ad hoc experts" should not provide transcripts of indistinct forensic audio, and a proposal for a better approach. *Crim. Law J.* 42, 298–302.
- Fry, H. (2021). What data can't do. *The New Yorker*. Available online at: <https://www.newyorker.com/magazine/2021/03/29/what-data-cant-do> (accessed June 26, 2022).
- Gibbons, J. (1995). "What got lost? The place of electronic recordings and interpreters in police interviews," in *Language in Evidence: Issues Confronting Aboriginal and Multicultural Australia*, ed D. Eades (Sydney: UNSW Press).
- Gilbert, D., and Heydon, G. (2021). Translated transcripts from covert recordings used for evidence in court: issues of reliability. *Front. Commun.* 6, 779227. doi: 10.3389/fcomm.2021.779227
- Gillon, G. (2007). *Phonological Awareness: From Research to Practice*. New York, NY: Guilford Press.
- González, R., Vásquez, V., and Mikkelsen, H. (2012). "Forensic transcription and translation" in *Fundamentals of Court Interpretation: Theory, Policy and Practice*, (Durham, NC: Carolina Academic Press), 965–1042.
- Grant, T. (2022). *The Idea of Progress in Forensic Authorship Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/9781108974714
- Green, J., Franquiz, M., and Dixon, C. (1997). The myth of the objective transcript: Transcribing as a situated act. *TESOL Quart.* 31, 172–176.
- Gurevich, O., Johnson, M., and Goldberg, A. (2010). Incidental verbatim memory for language. *Lang. Cognit.* 2, 45–78. doi: 10.1515/langcog.2010.003
- Hale, S., Goodman-Delahunty, J., and Martschuk, N. (2019). Interpreter performance in police interviews. *Differences between trained interpreters and untrained bilinguals. Interpret. Transl. Train.* 13, 107–131. doi: 10.1080/1750399X.2018.1541649
- Haworth, K. (2013). Audience design in the police interview: the interactional and judicial consequences of audience orientation. *Lang. Soc.* 42, 45–69. doi: 10.1017/S0047404512000899
- Haworth, K. (2018). Tapes, transcripts and trials. *Int. J. Evid. Proof.* 22, 428–450. doi: 10.1177/1365712718798656
- Hepburn, A., and Bolden, G. B. (2012). "The conversation analytic approach to transcription," in *The Handbook of Conversation Analysis*, eds J. Sidnell and T. Stivers (Oxford: Blackwell), 57–76. doi: 10.1002/9781118325001.ch4
- Heselwood, B. (2013). *Phonetic Transcription in Theory and Practice*. Edinburgh: Edinburgh University Press. doi: 10.1515/9780748691012
- Himmelmang, N. (2018). "Meeting the transcription challenge," in *Reflections on Language Documentation 20 Years After Himmelmang 1998*, eds B. McDonnell, A. Berez-Kroeker, and G. Holton (Honolulu: University of Hawaii Press), 33–40.
- Hoffman, D. (2019). *The Case Against Reality: Why Evolution Hid the Truth from Our Eyes*. (New York, NY/London: W. W. Norton and Company).
- JCCD (Judicial Council on Cultural Diversity) (2022). *Recommended National Standards for Working with Interpreters in Courts and Tribunals*, 2nd Edn. Available online at: <https://jccd.org.au/wp-content/uploads/2022/04/JCCD-Recommended-National-Standards-for-Working-with-Interpreters-in-Courts-and-Tribunals-second-edition.pdf> (accessed June 26, 2022).
- Jefferson, G. (2004). "Glossary of transcript symbols with an introduction," in *Conversation Analysis: Studies from the First Generation*, ed G. Lerner (Amsterdam: Benjamins), 13–31. doi: 10.1075/pbns.125.02jef
- Jenks, C. (2013). Working with transcripts: an abridged review of issues in transcription. *Lang. Linguist. Compass* 7, 251–261. doi: 10.1111/lnc3.12023
- Jones, T., Kalbfeld, J., Hancock, R., and Clark, R. (2019). Testifying while black. *Language* 95, e1–37. doi: 10.1353/lan.2019.0042
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar Straus Giroux.
- Kara, H. (2022). *Qualitative Research for Quantitative Researchers*. London: Sage.
- Knoch, U., and MacQueen, S. (2020). *Assessing English for Professional Purposes*. London/New York, NY: Routledge. doi: 10.4324/9780429340383
- Komter, M. (2019). *The Suspect's Statement: Talk and Text in the Criminal Process*. Cambridge: Cambridge University Press doi: 10.1017/9781107445062
- Linell, P. (1988). "The impact of literacy on the conception of language: the case of linguistics," in *The Written World*, ed R. Saljo (New York, NY: Springer), p. 41–58.
- Loakes, D. (2022). Does Automatic Speech Recognition (ASR) have a role in the transcription of indistinct covert recordings for forensic purposes? *Front. Commun.* 7, 803452. doi: 10.3389/fcomm.2022.803452
- Love, R. (2020). *Overcoming Challenges in Corpus Construction*. London/New York, NY: Routledge. doi: 10.4324/9780429429811
- Love, R., and Wright, D. (2021). Specifying challenges in transcribing covert recordings: implications for forensic transcription. *Front. Commun.* 6, 797448. doi: 10.3389/fcomm.2021.797448
- Mariottini, L. (2017). "Forensic interactions: power and (il)literacy in Spanish courtroom trials," in *Forensic Communication in Theory and Practice: A Study of Discourse Analysis and Transcription*, eds F. Orletti and L. Mariottini (Newcastle upon Tyne: Cambridge Scholars Publishing), p. 151–168.
- Munday, J. (2016). *Introducing Translation Studies: Theories and Applications, 4th Edn*. London/New York, NY: Routledge. doi: 10.4324/9781315691862
- NAJIT (National Association of Judiciary Interpreters and Translators). (2019). General guidelines and minimum requirements for transcript translation in legal settings. *NAJIT Position Papers Position Papers on Issues Affecting Court Interpreters and Translators*. Available online at: <https://najit.org/position-papers/> (accessed June 26, 2022).
- NSW Young Lawyers Criminal Law Committee. (2004). *Practitioner's Guide to Criminal Law*, 3rd Edn. Available online at: <https://crimlawcommittee.wordpress.com/practitioners-guide-table-of-contents/> (accessed June 26, 2022).
- Ochs, E. (1979). "Transcription as theory," in *Developmental Pragmatics*, eds E. Ochs and B. Schieffelin (New York: Academic Press), p. 43–71.
- Olson, D. (1994). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge: Cambridge University Press.
- Ong, W. (1982). *Orality and Literacy*. London: Methuen and co.
- Park, J., and Bucholtz, M. (2009). Introduction. *Public transcripts: entextualization and linguistic representation in institutional contexts. Text Talk Interdiscipl. J. Lang. Disc. Commun. Stud.* 29, 485–502. doi: 10.1515/TEXT.2009.026
- Pieraccini, R. (2012). *The Voice in the Machine: Building Computers that Understand Speech*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9072.001.0001
- Richardson, E., Haworth, K., and Deamer, F. (2022). For the record: questioning transcription processes in legal contexts. *Appl. Linguist.* 1–22. doi: 10.1093/applin/amac005. [Epub ahead of print].
- Ritchie, S. (2020). *Science Fictions: How Fraud, Bias, Negligence and Hype Undermine the Search for Truth*. New York, NY: Metropolitan Books.
- Roberts, A., and Ormerod, D. (2021). The full picture or too much information? Evidential use of body-worn camera recordings. *Crim. Law Rev.* 8, 620–641.
- Roberts, C. (1997). *Transcribing talk: issues of representation*. 31, 167–172.
- Scharf, H. (1989). The court reporter. *J. Legal History* 10, 191–227.
- Urban, G. (1996). "Entextualisation, Power and Replication," in *Natural Histories of Discourse*, eds M. Silverstein and G. Urban (Chicago: University of Chicago Press).

- Voutilainen, E. (2018). The regulation of linguistic quality in the official speech-to-text reports of the Finnish parliament. *CoMe Stud. Commun. Linguist. Cult. Med.* 2, 61–73.
- Wald, B. (1995). The problem of scholarly predisposition: G. Bailey, N. Maynor, & P. Cukor-Avila, eds., *The emergence of Black English: Text and commentary.* *Lang. Soc.* 24, 245–257.
- Walsh, M. (1995). “Tainted evidence”: literacy and traditional knowledge in an Aboriginal land claim,” in *Language in Evidence: Issues Confronting Aboriginal and Multicultural Australia*, ed D. Eades (Sydney: UNSW Press), p. 97–124.

Conflict of Interest: The author declares 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.

Copyright © 2022 Fraser. 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.