Transcribing screen-capture data: The process of developing a transcription system for multi-modal text-based data

Abstract

Transcription of audio data is widespread in qualitative research, with transcription of video-data also becoming common. Online data is now being collected using screen-capture or video software, which then needs transcribing. This paper draws together literature on transcription of spoken interaction and highlights key transcription principles, namely reflecting the methodological approach, readability, accessibility, usability. These principles provide a framework for developing a transcription system for multi-modal text-based data. The process of developing a transcription system for data from Facebook chat is described and reflected on. Key issues in the transcription of multi-modal text-based data are discussed, and examples provided of how these were overcome when developing the transcription system.

Keywords: Transcription, conversation analysis, screen-capture data, online interaction, text-based chat.

Introduction

A recent development in qualitative research has been the growing interest in online text-based data, such as online chats, e-mails, text messages, instant messaging interactions and so on (e.g., Baym, 2009; Coulson, 2005; Herring, 2007). While much of this data does not need to be transcribed, some online data is now being collected using screen-capture software (Bhatt & de Roock, 2013), which records actions represented on a computer screen. The use of these kinds of data poses challenges for transcription, which is an integral part of qualitative research. Transcription most commonly involves representing an oral language, with its attendant set of rules, as written language, with a different set of rules (Kvale, 1996). Previous literature has described the process of transcription and has provided 'how to' guides (e.g., Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993; Hepburn & Bolden, 2013; Jefferson, 2004). Transcription has also been critically examined, with many authors noting that it is not a neutral process, but rather is theory-laden (Bucholtz, 2000; Davidson, 2009; Lapadat & Lindsay, 1999; Ochs, 1979). While most of the literature has focused on transcribing spoken language, a growing literature discusses how to transcribe physical aspects of an interaction, such as body language, gaze, gesture and so on (e.g. Bezemer & Mavers, 2011; Goodwin, 1986; Heath, Hindmarsh, & Luff, 2010). This paper brings together the previous literature on transcription and transcription principles, and uses it to show how multi-modal text based data can be transcribed.

In the following section, I will provide an overview of the literature on transcription, including reviewing the development of multi-modal transcription for both spoken and online discourse. In the second section, I will identify the key principles which comprise the framework for developing a transcription system. The final sections provide examples of a transcription system developed for text-based screen-capture data. I will use these examples to demonstrate some of the challenges of transcribing multi-modal data, and how the

transcription principles guided the process.

Transcripts and transcription in qualitative research

Transcripts are used in qualitative research to aid analysis. A transcript is "compact, transportable and reproducible, and provides for easy random access unlike audio or video records" (Hepburn & Bolden, 2013, p. 75). In other words, paper transcripts can be accessed at any time, without needing a computer or any technical equipment. Transcripts allow data to be presented to others in journal articles, conferences, data sessions and so on. However, there is no 'gold standard' for how to transcribe spoken discourse and there is debate about how much, and what kind of, information to include in transcripts (Potter & Hepburn, 2005; Smith, Hollway, & Mishler, 2005). Potter and Hepburn argue that transcripts "should be transcribed to a level that allows interactional features to be appreciated even if interactional features are not the topic of study" (p.291). However, Smith, Hollway and Mishler suggest that this stance does not consider the differing aims of different types of qualitative research, and that for many approaches the inclusion of interactional features is irrelevant for the subsequent analysis. This debate encapsulates the issue that transcripts are not merely objective artefacts, but rather are theory-laden (Lapadat & Lindsay, 1998; Ochs, 1979). In other words, "data collection and transcription are affected by the theoretical interests of the analyst which inevitably determine which aspects of an interaction will be attended to and how they will be represented" (Jones, 2011, p. 9). Therefore, the finished transcript is not a simple record of the audio or video data, but instead is a record of the *approach* taken to the data by the transcriber (Bird, 2005; Lapadat & Lindsay, 1998).

The process of producing a transcript is often considered as the first stage of analysis (Edwards, 2003; Kvale, 1996; Lapadat & Lindsay, 1998, 1999), although it is also argued that even the initial viewing or hearing of the data requires the construction of meaning

(Ashmore, MacMillan, & Brown, 2004; Goodwin, 1994). The production of a transcript allows the researcher to familiarise themselves with the data, and subsequently identify interesting phenomena for analysis (ten Have, 2007). However, it is important that transcripts are not treated as data (Hutchby & Wooffitt, 2008) nor as an adequate substitute for watching or listening to the recorded data (Hepburn & Bolden, 2013; Heritage & Atkinson, 1984).

Multi-modal transcription

With the increased use of video in qualitative research, a number of systems for representing embodied conduct have been developed (e.g., Goodwin, 1980, 1986; Heath, et al., 2010). Bezemer and Mavers (2011) note that physical or bodily conduct is often simply described, such as 'shakes head'. Images are also used, including video stills, drawings and computer-generated images (e.g., Avital & Streeck, 2011; Goodwin, 2007; Heath & Luff, 2011), but their use varies considerably. Images are sometimes used alongside transcripts, highlighting particular actions or lines of dialogue (e.g., Craven & Potter, 2010). Others use images *as* the transcript, with relevant text overlaid on the images (e.g., Norris, 2004). However, due to the space which images require, these transcripts often only represent small sections of data (Bezemer, 2014). There are, of course, further transcription notations for multi-modal data in addition to those described above (Bezemer and Mavers, 2011), with some transcripts combining different methods. For example, Richardson and Stokoe (2014) use stills alongside transcripts, with relevant lines of dialogue imposed on the images as well as descriptions of actions in the transcript.

There are few examples of transcription of text-based multi-modal data. For studies of online data, authors sometimes choose to use screenshots to represent the actions occurring on-screen (e.g., Keating & Sunakawa, 2011), while others provide a written transcript (e.g., Beisswenger, 2008; Garcia & Jacobs, 1999). Beisswenger's transcript includes one column

for posted messages, another column for messages in construction, and a third column for embodied conduct taken from a video recording. Garcia and Jacobs, on the other hand, reproduce "what is visually available to each student on the screen" (p.343), and show what is appearing in the chat window, along with the message entry box for each participant. While these transcription systems are used to present data, it is rare that there is a description of *how* these systems were developed. This paper draws upon and develops previous literature on transcription by discussing the specific challenges of transcribing multi-modal text-based data.

Transcription principles

In this section, I will discuss a number of transcription principles drawn from the literature, which function as a framework for guiding the development of a transcription system. As previously discussed, transcription should not be seen an objective process, rather it is "a very complex process involving a series of interpretive judgements and decisions" (Müller & Damico, 2002, p. 300). The choices made during the transcription process are dependent upon the questions the researcher asks of that data and their methodological and theoretical approach (Lapadat & Lindsay, 1998; Ochs, 1979). In other words, a key principle is that the transcript will, and should, inevitably reflect the analytic method and research question.

Another principle is that a transcript should be readable. As Ochs (1979) argues "a transcript that is too detailed is difficult to follow and assess. A more useful transcript is a more selective one" (p44). Therefore, a transcript will necessarily be selective, and so transcribers will need to consider what should be included and excluded, based upon the goals of the research (Davidson, 2009; Lapadat, 2000; Müller & Damico, 2002).

A third, and potentially competing principle, is that the transcript must be useful to the researcher. Therefore, it must include *enough* information so that it can be used for current,

and potentially future, analytic interests (Du Bois, 1991). However, a balance must be struck between the transcript being usable and readable.

A fourth principle is that the transcript should be accessible to others in similar fields of research, by using a similar layout or 'borrowing' symbols from other well-established systems (Du Bois, 1991) . While readability and usability relate to the researcher's use of the transcript, accessibility focuses on whether others are able to use and understand it.

These principles are not rigid rules, but rather are a framework for guiding the process of both developing a transcription system and subsequently producing transcripts. In the rest of this paper I will discuss the process of developing a transcription system specifically for multi-modal text based data. I will show how the principles discussed here are relevant for managing the challenges of transcribing these kinds of data. First, though, I will provide a brief overview of the data collected and methodological approach taken to analysis, because – as discussed above – such factors impact upon the production of a transcript.

Data and methodological approach

The data comprised a corpus of screen-capture videos of Facebook chats. Facebook chat is a text-based instant messaging service, available through the social networking site Facebook, allowing users to interact with their friends in real time[1]. Four participants downloaded screen-capture software on to their computers, and were asked to engage in Facebook chats as normal, but to record their screens when doing so. All participants gave informed consent to have their chats recorded, and they also gained consent from their chat partners. 47 screen-capture chats were collected comprising around 25 hours of recordings.

[1] At the time of data collection, Facebook chat was a quasi-synchronous means of interacting with others online. After data collection was completed, Facebook combined the asynchronous private messaging and quasi-synchronous chat features.

The aim of the research was to investigate how instant messaging interaction is organised, using conversation analysis (henceforth CA) to analyse the data. CA examines instances of naturally-occurring interaction to analyse how social action can be seen as patterned and orderly (Heritage & Atkinson, 1984). The key findings of CA are that turns-at-talk incorporate actions, in other words, they are 'doing things' such as inviting, requesting, offering and so on (Drew, 2005). In addition, CA finds that talk is organised sequentially, so one action will project a particular next action. CA is also concerned with how participants mutually co-ordinate turn-taking in conversation (Schegloff, 2007). The broad analytic interests of CA are therefore action, sequences and turn-taking, and this is reflected in the types of transcripts produced.

CA transcripts are designed to include details of not only what is said, but also details of how a turn is delivered, such as the pitch, volume, speed or prosody of a turn (Hepburn & Bolden, 2013). The most common method of transcription in CA is the Jefferson (2004) system, which incorporates ways of representing temporal features, utterance alignment, speech delivery and intonation (Hepburn & Bolden, 2013; Roberts & Robinson, 2004). The Jefferson system is based largely on notation which is familiar from written interaction, including capital letters for volume, underlining for emphasis and so on (Hepburn & Bolden, 2013).

For the Facebook chat data, a transcription system was developed which would reflect my research interests and which was accessible, as far as possible, to other CA researchers. There were three key features which distinguished this data from spoken interaction, and which posed challenges for transcription: firstly, the interaction itself was text-based, as opposed to representing spoken language as text; Secondly, some data was only available via the screen-capture and was not available to both participants in the chat; and thirdly, the on-screen data often involved moving text (for example, writing or deleting messages), and this

needed to be accurately represented and distinguished from the chat itself. In the following sections I will discuss the process of developing a transcription system, and show how these issues were addressed.

Designing a multi-modal transcription system

In this section I will outline some initial decisions made in the process of developing the transcription system. I will then discuss the layout and some of the symbols used throughout the transcript. Finally, I will show how overlaps, writing and deleting are represented in the transcript.

The initial decision-making process

Hammersley (2010) notes that a number of decisions need to be made during transcription, including what to transcribe, how much to transcribe and how much detail to include. Here, I will discuss the decisions involved in the initial stages of transcribing screen-capture data. The first step in developing the transcription system involved watching the screen-capture videos, and noting relevant details. This initial stage functioned as a 'noticing device' (ten Have, 2007, p. 95), allowing me to identify interesting phenomena for analysis and was, effectively, the first stage of analysis (Edwards, 2003). In text-based data, the main aspect to be included is the actual interaction, in this case, the text of the Facebook chat. The screen-capture data provided additional information, including message construction, overlapping writing, using other websites, and so on, which was only available to the participant recording their screen.

There were many features in the video data which could be included, but based on the principle of readability it was necessary to consider what features would be analytically relevant, and to exclude those which did not appear to be. For example, I chose not to include

the mouse movement as this action would not assist in answering the research question. In addition, the inclusion of mouse movements in multi-modal transcripts often makes the transcript more complex (Laurier, Forthcoming), and thus less readable.

Another decision related to how much detail to include. References to activities outside of the chat were included, but with fairly limited information, as demonstrated in the Figure 1.

```
1   K*:     💻 6.0 University e-mail 💻
2           ((Opens Microsoft Word))
```

Figure 1. Extract from Facebook chat.

K spends 6 seconds using her university e-mail (line 1), before opening word processing software (line 2). These lines of text represent images but the action is described rather than shown. The description of these activities is brief; there is no detail such as whether K was writing, reading or organising her e-mail for example, as this was deemed irrelevant for analysis. The availability of the full screen-capture video also raised another issue in terms of selectivity. Participants often engaged in more than one chat simultaneously and therefore the decision needed to be made whether to include all chats in one single transcript or to represent each single interaction in its own transcript. Including all the interactions would represent the participants' lived experience of Facebook chat (Meredith & Potter, 2014), but the aim of the research was to focus on one-to-one interaction. Therefore, guided by the principle that the transcript should reflect the aims of the research (Lapadat, 2000; Ochs, 1979), I transcribed each single interaction separately. An additional consideration was that a transcript of the entire screen would most likely have been extremely complex and therefore both unreadable and unusable for analysis (Du Bois, 1991).

One final decision at this stage was whether to use transcription software such as Transana. I chose not to do this for a number of reasons. Firstly, in keeping with the principle of accessibility (Du Bois, 1991), I wanted to produce the transcript as Jefferson transcripts are produced, where Microsoft Word is most commonly used (Heritage, n.d.). Secondly, not using transcription software allowed for work on the transcript on any computer, without needing to download additional software, making it more practical.

So far, I have laid out the basic decision-making processes of developing a transcription system. Some of these decisions are similar to those made when transcribing spoken interaction, but some issues do relate to the data being transcribed. In the following section, I will discuss some of the more practical decisions about the transcript, related to the layout and symbols.

Layout and symbols of the screen-capture transcript

One key decision was whether the transcript should be in vertical, column or partiture format (Edwards, 2003). Most spoken transcripts are "arranged in the conventional 'play-script' layout" (Jones, 2011, p. 14), meaning they are read from top-to-bottom and left-to-right. While some transcripts for spoken interaction use a column-based format, this can give "the impression (due to left-to-right reading bias) that the speaker whose utterances are leftmost is the more dominant in the interaction" (Edwards, 2003, p. 326; see also Ochs, 1979). A similar issue could arise with multi-modal transcripts, as one part of the interaction – the chat or on-screen actions – could be given precedence in a column-based transcript. One consideration for the Facebook chat transcript was that the Jefferson system predominantly uses a vertical format, although it borrows occasionally from the partiture format (ten Have, 2007). Therefore, using a vertical format would make the transcript more accessible to other conversation analysts (Du Bois, 1991). An example of the layout of the final transcript is

shown in Figure 2.

1	2	3	4	5
1	2.17	0.15	Isla:	wait wait wait
2				(2.0)
3			I*:	✍ i ~~i~~ ‖ (2.0) ‖ ✍
4	2.23	0.06	Callum:	‖ new ‖

Figure 2. Extract from Facebook chat.

There are columns in the transcript, which are numbered for the purposes of this example. The interaction itself is in a single column (Column 5), and therefore resembles the vertical format, with both the on-screen activities and the chat in the same column; this decision was made for three reasons. Firstly, breaking up the on-screen activities and the chat could mean that, in the course of the analysis, the chat was given precedence. Secondly, the use of columns breaks up the linearity of the interaction. Thirdly, using columns makes the transcript less readable and useable for conversation analysis, as it is difficult to accurately indicate the occurrence of overlaps, which are key for the analysis of turn-taking. Therefore, in the screen-capture transcript, both the actions taken from the screen-capture and the chat are included in a single column.

Columns are used in this transcript for indicating line numbers (Column 1) and participant identification (Column 4), as in a Jefferson transcript. In Jefferson transcripts, time between and within turns is timed to tenths of seconds and placed within the interaction (Hepburn & Bolden, 2013). However, in the screen-capture transcript, due to the nature of textual interaction, there are different timings to be presented. Firstly, time between sent turns is relevant, and available, for both participants in the chat and these are represented in column 3 (15 seconds; 6 seconds). In addition, the cumulative time elapsed is indicated in

column 2 (2 minutes 17 seconds; 2 minutes 23 seconds). Secondly, there are on-screen gaps, taken from the screen-capture, when nothing is happening on-screen for the participant recording the data (line 2). Thirdly, there are pauses in the construction of messages, where the writer momentarily stops writing (line 3). These latter two timings are only available from the screen-capture, and these are presented within the interaction itself (Column 5). The other timings, available to both participants, are placed alongside the interaction, as a method of distinguishing between the different timings.

In Column 4 the full (anonymised) name of the participant is used in lines 1 and 4, whereas only the initial of the participant is used in line 3. This distinction relates to how different data types can be contrasted "so that readers of a transcription will know at every moment what kind of information they are taking in" (Du Bois, 1991, p. 79). In this case, the distinction is between the visible interaction available to both participants, and the on-screen actions, available only to one participant. When a line refers to a part of the visible interaction, the participants' full names are used. When only an initial with an asterisk is used, as in line 3, the information is taken directly from the screen-capture and refers to some action occurring on-screen.

Another way of contrasting data types was through highlighting the turns which were part of the interaction and so visible to both participants (lines 1 and 4). This means that for a reader who is not familiar with the transcript, it is clear which parts of the chat are visible to both participants, thus making it more accessible and readable (Du Bois, 1991).

Figure 2 also shows some of the symbols used throughout the transcript. When deciding which symbols to use, accessibility was the key principle abided by and conventions from the Jefferson system were used where possible. However, while the Jefferson system uses some notations from written interaction, such as capital letters, underlining and punctuation (Hepburn & Bolden, 2013), this was not possible in text-based chat as these

written notations were used as part of the interaction. Therefore, symbols needed to be chosen which were not commonly used in written language, but which were widely available (Du Bois, 1991). In the Facebook chat transcript, symbols were chosen from the 'Wingdings' and 'Zingbats' fonts on Microsoft Word. The symbols were chosen because they best represented the actions on-screen, but also because they seemed unlikely to be used in everyday written interaction.

Writing and overlaps

In Facebook chat, as with most other instant messaging services, the construction and sending of messages are separate, so message construction is not visible to the recipient. Therefore, the issue arises when transcribing text-based screen-capture data of how to distinguish between text which is not visible to the co-participant, and text which is visible to both participants. While some methods have been described above, such as highlighting turns, the following examples show in more detail how writing is represented.

1			I*:	✍ *wait wait wait* ✍
2	2.17	0.15	Isla:	wait wait wait

Figure 3. Extract showing writing in Facebook chat.

In Figure 3 the use of an initial in line 1, rather than a full name, indicates that this action occurs on-screen. In contrast, in line 2, which is visible to both participants, the full name is used and the line is highlighted. The writing symbol (✍) is used in line 1 to show that Isla is constructing a message, and it is also placed at the end of message construction, which is particularly important when it occurs over a number of lines of the transcript. Du Bois (1991) suggests that symbols should be 'iconic', that is, they should have some link to the action

they are designed to represent. The writing symbol is 'iconic', as it visually represents the action it is symbolising. It is also readily available through Microsoft Word and was not available in Facebook chat, so would be less likely to be used during an ordinary chat. The message being constructed is written in italics, to distinguish it from messages sent to the chat, thus aiding readability and usability.

From the screen-capture video, the overlap of message construction and sending was visible (as seen briefly in Figure 2) and these phenomena were included in the transcript, as they were relevant to the research question. The following example shows how overlap is represented.

1			I*:	✍ *yeah i need* ‖ *to* ‖ ↔
2	06:36	00:21	Joe:	‖ some one tried to add
3				photos of me from friday ☺ ‖
4			I*:	↔ *cook* ✍
5	06:37	00:01	Isla:	yeah I need to cook
6				(2.0)
7			I*:	✍ *and i*(2.0)[*think we might watch* ↔
8			J*:	[✍ *writing* ↔
9			I*:	↔ *a watch a film tod ~~d~~ o* ✍]
10			J*:	↔ *writing*] ↔
11	06:48	00:11	Isla:	and i think we might watch a film
12				too
13			J*:	↔ 1.0 *writing* ✍
14	06:49	00:01	Joe:	yea you need cock? ….....what?
15			I*:	✍ *who>>* ✍
16	06:50	00:01	Isla:	who>>

Figure 4. Transcript showing overlapping writing.

[Figure 5 about here]

In Figure 4 the content of Isla's message construction is available, whereas Joe's is not, as Isla is the participant recording her screen. It is possible to see that Joe is writing, as a small writing symbol appears in the corner of the chat window (see Figure 5). Therefore, from viewing the screen-capture video it is possible to see when both parties are constructing

messages. As with the Jefferson system, overlaps are indicated using square brackets (lines 7 and 8), making it more accessible to conversation analysts.

Joe's message at lines 2 and 3 is sent while Isla is constructing hers, and in line 4 Isla continues to write her next turn. In the transcript this is indicated by the double headed arrow ('↔'), representing 'latching', placed at the end of line 1 and the start of line 4. This symbol indicates that the writing is continuous. From lines 7 to 10, writing symbols, overlapping brackets and latching symbols are used to indicate that the two parties are writing simultaneously.

In line 1 Isla starts to construct the turn that eventually appears at line 5; an action which is only available for Isla. At the end of the line, while Isla is writing 'to', Joe sends a message to the chat (lines 2 and 3), and this is represented by the use of ‖ symbols placed around the parts that occur at the same time. The ‖ symbols are placed around the entire message which occurs simultaneously, even if this runs over more than one line. Here, then, the partiture format is used for actions occurring at the same time, similarly to Jefferson transcripts, thus making the transcript more accessible. The features highlighted here show how the text-chat and multi-modal features can be represented in the same transcript using symbols and descriptions, rather than images. The overlap of on-screen actions and text-chat demonstrates how including the interaction in a single column makes it clearer how overlap occurs.

The production of such a detailed transcript reflects the aims of the research, which were to examine the sequential and turn-taking aspects of instant messaging conversations. For example, in Figure 5 Joe's humourous message to Isla at line 14 ('yea you need cock….what?') appears to be unrelated to the previous turn. However, by transcribing the overlapping writing, we can see that this sequential disruption occurs because of the overlapping writing and posting. Isla posts the first-pair part (Schegloff & Sacks, 1973) to

Joe's message ('yeah I need to cook') in line 5. Isla then continues to construct another message in line 7. What is apparent from the transcript is that Joe starts to construct a message in line 8, while Isla is still writing her subsequent message. Isla finishes constructing her message a second prior to Joe and therefore her message appears first. So, we can see, firstly, how turn-taking and sequence organization are impacted by the ability to write messages simultaneously (see Meredith, forthcoming). Secondly, the fact that simultaneous writing is not accountable, as simultaneous speech might be (see Schegloff, 2000) means that we can start to understand some of the interactional norms of instant messaging interaction.

Deleting and editing messages

As the Facebook chat interaction is text-based and message construction occurs separately from sending, participants are able to edit their messages, which raises the question of how to accurately transcribe this editing. In the transcripts, strikethrough of letters was used to represent deletion, utilising methods from other forms of written communication (Du Bois, 1991; Hepburn & Bolden, 2013). However, due to the issues of representing moving text in a static format, the representation of deletion was slightly more complex than first anticipated. Consider, for example, Figure 6 which shows a message construction and deletion.

1		I*:	✍ ~~holy (.) shit (.)t st~~
2			~~hit joe - whereve e were you~~
3			~~lastnight night ? ? and i bet~~
4			~~you can't remember a thinkg glo~~
5			~~g g lol (3.0) i th (1.0)~~
6			(2.0) *oh dar* ✍
7	01.23	Isla:	oh dar

Figure 6. Transcript of deletion in Facebook chat.

In this example, it is difficult to see what the deleted message was, as all the letters are struck-through. As will be shown below, the detail of what is being deleted can be relevant for analysis. In addition, any nuances of message construction are lost because the entire message is deleted. Therefore, in order to enhance the usability of the transcript, the decision to re-write the letters was taken, as shown in the following example.

1		I*:	✍ holy (.) shit ~~tih~~ (.) ~~s~~ t ~~t~~ st ~~t~~
2			hit joe – wherewe ~~ewe~~ e were you
3			lastnight ~~thgin~~ night ? ~~–?~~ ? and i bet
4			you can't remember a thinkg ~~gk~~ glo ~~–olg~~
5			g ~~–g~~ g lol (3.0) i th ~~ht i~~ (1.0)
6			~~lol gniht a rebmemer t'nac uoy teb I dna ?thgin~~
7			~~tsal uoy erew erehw – eoj this yloh~~
8			(2.0) oh dar ✍
9	01.23	Isla:	oh dar

Figure 7. Revised transcript of deletion.

In Figure 7, it is easier to see how the message was constructed, moment-by-moment. The letters which are being deleted are re-written in the order in which they are deleted (in reverse). For minor corrections, this representation of the deleted letters is mostly unproblematic. However, for the major deletion in this extract, it is difficult to see – at first glance – what has been deleted. Therefore, it was decided that the letters would be re-written as normal, thus enhancing the usability and accessibility of the transcripts. The following example shows the final version of the transcript.

1	I*:	✍ holy (.) shit ~~hit~~ (.) ~~s~~ t ~~t~~ st ~~t~~
2		hit joe – wherewe ~~ewe~~ e were you
3		lastnight ~~night~~ night ? ~~–?~~ ? and i bet
4		you can't remember a thinkg ~~kg~~ glo ~~–glo~~

5				*g ~~g~~ g lol (3.0) i th ~~i th~~ (1.0)*
6				~~*holy shit joe - where were you last*~~
7				~~*night? and i bet you can't remember*~~
8				~~*a thing lol*~~ (2.0) *oh dar* ✍
9	01.23	Isla:		oh dar

Figure 8. Final transcription of deletion.

The transcript appears more readable and usable when written in the format in Figure 8. It is possible to see the construction of the message moment-by-moment, but also to see clearly the deleted text in lines 6-8. Having a clear representation of deletions enabled the analysis of how participants orient to the potential implications of doing a particular action in the conversation (Drew, Walker, & Ogden, 2013). The deletion in Figure 8 is significant because it represents a shift from Isla asking further questions about the topic to merely assessing the situation. Initially, Isla issues an inquiry 'where were you last night?'. The second action is an assertion 'I bet you can't remember a thing lol', which could project a confirmation or denial, or perhaps a humorous account. This completed message would have projected a further telling from Joe about his evening. However, in her eventual turn Isla does not close down the topic explicitly, but neither does she invite further talk on it (See Meredith & Stokoe, 2014).

Discussion

This paper has demonstrated how a framework of transcription principles can be used for developing a transcription system for text-based multi-modal data. Previous literature has shown how such data can be transcribed (e.g., Beisswenger, 2008; Garcia & Jacobs, 1999), but transcripts are often presented with little explanation of the choices made (Davidson, 2009). As Lapadat (2000) suggests "transcription decisions and processes employed during

data collection and analysis need to be explained clearly and thoroughly in the write-up" (p.217). At the outset it was suggested that there were three key challenges which multi-modal text-based data posed: 1) the interaction was text-based; 2) some data was only available via screen-capture and not to both participants; and 3) the on-screen data involved moving text. The transcript presented here demonstrates how these particular challenges can be overcome. This paper therefore contributes to the literature on multi-modal transcription by offering an in-depth explanation of how a transcription system was developed.

This transcript is, as with all transcripts, a record of the approach taken to the data (Bird, 2005; Lapadat & Lindsay, 1998), that is, conversation analysis. Consequently, the debates around transcription in spoken interaction (Potter & Hepburn, 2005; Smith, Hollway & Mishler, 2005) are also relevant here. Potter & Hepburn argue that all interview extracts "should be transcribed to a level that allows interactional features to be appreciated even if interactional features are not the topic of study" (p.291). A similar argument could be applied to multi-modal transcripts; that is, that details of overlaps, writing and editing should be included in all transcripts. However, depending on the focus of the analysis, this level of detail may not be necessary. Considering that there can be no 'neutral' transcription (Bucholtz, 2000; Lapadat & Lindsay, 1998), ensuring that the transcript itself reflects the aims of the research is perhaps most important when developing a transcription system. The researcher still has to make decisions about what to include and exclude from the transcript, principally based on how relevant certain details are to the research question. As a result, the development of a transcription system can be seen as a key part of the analysis (Edwards, 2003). For conversation analysts, features such as overlap, deletion and editing are important for the analysis, as demonstrated by the presentation of some brief analysis in this paper.

The four transcription principles were developed based on spoken interaction (e.g., Du Bois, 1991; Lapadat, 2000; Ochs, 1979). However, it is important to note that these

principles do not encapsulate all those discussed throughout the literature. For example, Edwards (2003) includes principles around computational tractability and visual display, and Du Bois' work (1991) discusses robustness and economy. However, such transcription principles are seemingly specific to transcribing spoken interaction. Therefore, the four principles used in this paper are chosen because they can be applied to a broader range of data. There may be questions over the extent to which the transcript is in fact readable or accessible to other conversation analysts. I have used a similar format to the Jefferson system, including borrowing some aspects of the partiture format, and some transcription symbols; however, it clearly deviates quite significantly from a Jefferson transcript. Therefore, for newcomers, it may be that this transcript is not particularly accessible or readable. However, for many new to Jefferson transcripts, time must be spent learning to read and do this form of transcription (ten Have, 2007). There are, for example, chapters in many introductory texts which cover Jefferson transcription (e.g., Hutchby & Wooffitt, 2008; ten Have, 2007) as well as a number of online tutorials providing guidance for new CA scholars (e.g., Antaki, 2011; Scheloff, n.d.). By including interactional features, both in text-based multi-modal and spoken transcripts, time is required to learn how to work effectively with the transcript, thus impacting upon its accessibility. It is notable that even when presenting the transcript in this paper, I used an image to demonstrate a feature of the technology. It is, therefore, certainly not my argument that the use of images is redundant when presenting multi-modal data. However, when analysing the data it is useful to have a transcript for noting observations or analytic comments (ten Have, 2007).

One potential issue with the data used in this paper is that having access to only one participant's screen may have implications for the accuracy of the transcript. For example, a message sent from one participant may not appear on the co-participant's screen immediately, due to 'lag' (Herring, 1999). By only having one participant's screen recorded,

it is not possible to examine the extent to which lag occurs and to reflect that in the transcript. Another related issue is that the detail of message construction can only be transcribed for one participant. The writing symbol appears in the chat window to indicate that the co-participant is constructing a message, but it is not possible to indicate whether they are writing, deleting or editing messages or whether, perhaps, the writing symbol has merely appeared in error. It is important to remember, though, that this also reflects the experience of the participant recording the screen, as they also only have access to their side of the chat. However, if screen-capture data were to be collected from both parties in the chat, the transcript could reflect the full chat, rather than one participant's side of it, although it would be particularly important to ensure that the transcript was still readable and usable when representing both participants' actions.

In contrast to spoken interaction, there is significant variability in the types of online interaction which could be recorded (see Herring, 2007 for a taxonomy of online communication). For data from online sites such as Instagram or Tumblr, which incorporate images, or sites such as Twitter, which involve far greater numbers of participants, the transcription system presented is likely to be unsuitable. However, the transcription principles laid out in this paper would be able to guide the development of transcription systems for different types of online data.

For conversation analysts, this paper demonstrates that text-based multi-modal interaction can be a rich source of interactional data. By collecting screen-capture data, interaction can be recorded in 'real-time', and this paper shows that it is possible to present these data in a readable transcript. Such findings should provide incentive to conversation analysts to continue to work with these kinds of data. For those working with multi-modal online data, this paper has provided some insight in to how such data can be presented, which does not rely on screen-shots. As new methods for collecting online data develop, it will be

important for researchers to consider the development of methods of transcribing and presenting such data as an integral part of the research and analysis (Edwards, 2003).

Conclusion

The aim of this paper was to demonstrate how text-based multi-modal data can be transcribed using a framework of transcription principles. The framework was based on four principles of transcription - accessibility, usability, readability and reflecting the aims of the research. There are three key challenges for transcribing text-based video data: 1) managing the text-based interaction; 2) representing text which is available to either one or both participants; and 3) representing 'moving' text. This article has argued that it is possible to overcome those challenges, through developing a transcription system according to the four basic principles of transcription. A transcription system, such as the one described in this article, can allow for a clear analysis of both text and video data. It is possible that such a framework could be used for developing transcription systems in the future.

Funding

This work was supported by the Economic and Social Research Council [Grant number: ES/I903321/1]

Acknowledgements

I would like to thank Elizabeth Stokoe and Charles Antaki for their advice at various stages of writing this article.

References

Antaki, C. (2011). CA tutorial: Transcription. From http://homepages.lboro.ac.uk/~ssca1/transintro1.htm [date accessed 23/03/2015]

Ashmore, M., MacMillan, K., & Brown, S. D. (2004). It's a scream: Professional hearing and tape fetishism. *Journal of Pragmatics, 36*, 349-374.

Avital, S., & Streeck, J. (2011). Terra incognita: Social interaction among blind children. In J. Streeck, C. Goodwin & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 169-181). Cambridge: Cambridge University Press.

Baym, N. K. (2009). The emergence of on-line community. In S. G. Jones (Ed.), *Cybersociety 2.0: Revisiting computer-mediated communication and community* (pp. 35-68). London: Sage.

Beisswenger, M. (2008). Situated chat analysis as a window to the user's perspective: Aspects of temporal and sequential organization. *5*(Article 6). Retrieved from http://www.languageatinternet.org/articles/2008/1532

Bezemer, J. (2014). Multi-modal transcription: A case study. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts: A reader in multimodality* (pp.155-170). Berlin: Mouton de Gruyter.

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: a social semiotic perspective. *International Journal of Social Research Methodology, 14*(3), 191-206.

Bhatt, I., & de Roock, R. (2013). Capturing the sociomateriality of digital literacy events. *Research in learning technology, 21*. Available at: <http://www.researchinlearningtechnology.net/index.php/rlt/article/view/21281>. doi:http://dx.doi.org/10.3402/rlt.v21.21281.

Bird, C. M. (2005). How I stopped dreading and learned to love transcription. *Qualitative Inquiry, 11*(2), 226-248.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics, 32*, 1439-1465.

Coulson, N. S. (2005). Receiving social support online: An analysis of a computer-mediated support group for individuals living with irritable bowel syndrome. *CyberPsychology & Behavior, 8*(6), 580-584.

Craven, A., & Potter, J. (2010). Directives: Entitlement and contingency in action. *Discourse Studies, 12*(4), 419-442.

Davidson, C. (2009). Transcription: Imperatives for qualitative research. *International Journal of Qualitative Methods, 8*(2), 35-52.

Drew, P. (2005). Conversation Analysis. In K. L. Fitch & R. E. Sanders (Eds.), *Handbook of language and social interaction* (pp. 71-102). Mahwah, N.J: Lawrence Erlbaum.

Drew, P., Walker, T., & Ogden, R. (2013). Self-repair and action construction. In M. Hayashi, G. Raymond & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 71-94). Cambridge: Cambridge University Press.

Du Bois, J. W. (1991). Transcription design principles for spoken discourse research. *Pragmatics, 1*, 71-106.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45-90). New Jersey: Lawrence Erlbaum.

Edwards, J. (2003). The transcription of discourse. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp.321-348). Malden, Massachusetts: Blackwell Publishers.

Garcia, A. C., & Jacobs, J. B. (1999). The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on Language and Social Interaction, 32*(4), 337-367.

Goodwin, C. (1980). Restarts, pauses and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry, 50*, 372-302.

Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. *Semiotica, 62*(1/2), 29-49.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.

Goodwin, C. (2007). Participation, stance and affect in the organization of activities. *Discourse & Society, 18*(1), 53-73.

Hammersley, M. (2010). Reproducing or constructing? Some questions about transcription in social research. *Qualitative Research, 10*(5), 553-569.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. London: Sage publications.

Heath, C., & Luff, P. (2011). Gesture and institutional interaction. In J. Streeck, C. Goodwin & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 276-288). Cambridge: Cambridge University Press.

Hepburn, A., & Bolden, G. B. (2013). Transcription. In T. Stivers & J. Sidnell (Eds.), *The Handbook of Conversation Analysis* (pp. 57-76). Oxford: Wiley-Blackwell.

Heritage, J. (n.d.). Advice on setting up a word document for transcription Retrieved 24th July 2014, from http://www.paultenhave.nl/Transcription%20-%20WORD.html

Heritage, J., & Atkinson, J. M. (1984). Introduction. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 1-15). Cambridge: Cambridge University Press.

Herring, S. C. (1999). Interactional coherence in CMC. *Journal of Computer-Mediated Communication, 4*(4). Retrieved from http://onlinelibrary.wiley.com/enhanced/doi/10.1111/j.1083-6101.1999.tb00106.x/

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet, 4*, Article 1. Retrieved from http://www.languageatinternet.org/articles/2007/761

Hutchby, I., & Wooffitt, R. (2008). *Conversation Analysis: Principles, practices and applications* (2nd ed.). Cambridge: Polity Press.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-23). Philadelphia: John Benjamins.

Jones, R., H. (2011). Data collection and transcription in discourse analysis: A technological history. In K. Hyland & B. Paltridge (Eds.), *The continuum companion to discourse analysis* (pp. 9-21). London: Continuum International.

Keating, E., & Sunakawa, C. (2011). "A full inspiration tray": Multimodality across real and virtual spaces. In J. Streeck, C. Goodwin & C. LeBaron (Eds.), *Embodied interaction:*

Language and body in the material world (pp. 194-204). Cambridge: Cambridge University Press.

Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. London: Sage.

Lapadat, J. C. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology, 3*(3), 203-219.

Lapadat, J. C., & Lindsay, A. C. (1998, 13-17 April). *Examining transcription: A theory-laden methodology.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Lapadat, J. C., & Lindsay, A. C. (1999). Transcription in research and practice: From standardization of technique to interpretive positionings. *Qualitative Inquiry, 5*(1), 64-68.

Laurier, E. (Forthcoming). Captures and blow-ups: On finally being able to read video recordings.

Meredith, J., & Potter, J. (2013). Conversation analysis and electronic interactions: Methodological, analytic and technical considerations. In H. Lim & F. Sudweeks (eds.), *Innovative methods and technologies for electronic discourse analysis* (pp.370-393) IGI Global, Hershey, USA.

Meredith, J., & Stokoe, E. (2014). Repair: Comparing Facebook 'chat' with spoken interaction. *Discourse & Communication, 8*(2), pp.181-207.

Meredith, J. (Forthcoming). Affordances and practices: The use of screen-capture data to understand disrupted turn adjacency in instant messaging interaction. *Manuscript in preparation.*

Müller, N., & Damico, J. S. (2002). A transcription toolkit: theoretical and clinical considerations. *Clinical Linguistics & Phonetics, 16*(5), 299-316.

Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. London: Routledge.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schieffelin (Eds.), *Developmental Pragmatics*. London: Academic Press.

Potter, J., & Hepburn, A. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology, 2*(4), 281-307.

Richardson, E., & Stokoe, E. (2014). The order of ordering: Requests, objects and embodied conduct in a public bar. In M. Nevile, P. Haddington, T. Heineman & M. Rauniomaa (Eds.), *Interacting with objects: Language, materiality, and social activity* (pp. 31-56). Amsterdam: John Benjamins.

Roberts, F., & Robinson, J. D. (2004). Interobserver agreement on first-stage conversation analytic transcription. *Human Communication Research, 30*(3), 376-410.

Schegloff, E.A. (n.d.). Transcription module. From http://www.sscnet.ucla.edu/soc/faculty/schegloff/TranscriptionProject/ [date accessed 23/03/2015)

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society, 29*, 1-63.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University press.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*(4), 289-327.

Smith, J. A., Hollway, W., & Mishler, E. G. (2005). Commentaries on Potter and Hepburn,'Qualitative interviews in psychology: problems and possibilities'. *Qualitative Research in Psychology, 2*(4), 309-325.

ten Have, P. (2007). *Doing Conversation Analysis: A practical guide* (2nd ed.). London: Sage Publications