# *Data collection and transcription in discourse analysis: a technological history*

**Data Collection and Transcription in Discourse Analysis:**
**A Technological History**

Rodney H. Jones
University of Reading

Data collection as mediated action

Since the publication of Elinor Ochs's groundbreaking 1979 article 'Transcription as Theory', it has become axiomatic that 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 (see also Edwards 1993, Mishler 1991). Since then, much of the debate around transcription has focused on choosing the 'best system' for transcribing spoken discourse (see for example DuBois et al. 1993, Psathas and Anderson, 1990) or 'multimodal interaction' (Baldry and Thibault 2006, Norris 2004) in order to serve the theoretical demands of particular approaches to discourse, or arguing about the need for standardization in transcription conventions (Bucholtz 2007, Lapadat and Lindsay 1999). In order to productively engage in such debates, however, it is necessary to consider more practical questions about data collection and transcription having to do with the *materiality* of what we call data and the effects of the technologies we use to collect and transcribe it on the ways we are able to formulate theories about discourse in the first place.

The focus of this chapter will be less on narrow questions about the best way to collect and transcribe data and more on data collection and analysis as cultural and material practices of discourse analysts (Jaffe 2007). In particular I will focus on how, over the past half century, these practices have been affected by different technologies such as tape-recorders, video cameras and computers, each of which made new kinds

of knowledge and new kinds of disciplinary identities possible, and each of which fundamentally changed our understanding of discourse itself.

The theoretical framework I will use to be approach these issues is mediated discourse analysis (Norris and Jones 2005,). Central to this perspective is the concept of mediation, the idea that all (inter)actions are mediated through cultural tools (which include technological tools like tape recorders and semiotic tools like transcription systems), and that the affordances and constraints of these tools help to determine what kinds of actions are possible in different circumstances. This focus on mediation invites us to look at data collection and transcription as physical actions which take place within a material world governed by a host of technological, semiotic and sociological affordances and constraints on what can be captured from the complex stream of phenomena we call 'social interaction', what can be known about it, and how we as analysts exist in relation to it, affordances and constraints that change as new cultural tools are introduced.

Mediated discourse analysis allows us to consider data collection and transcription as both *situated practices*, tied to particular times, places and material configurations of cultural tools, and *community practices*, tied to particular disciplinary identities.

Five processes of entextualization

Nearly all of the disciplinary practices discourse analysts engage in involve 'entextualization' -- transforming actions into texts and texts into actions. We turn ideas into research proposals, proposals into practices of interviewing, observation and recording, recordings into transcripts, transcripts into analyses, analyses into academic papers and academic papers into job promotions and academic accolades. Ashmore and Reed (2000) argue that the business of an analyst consists chiefly of

creating *artifacts*—such as transcripts and articles—that are endowed with both 'analytic utility' and professional value.

Bauman and Briggs 1990 define 'entextualization' as the process whereby language becomes detachable from its original context of production and reified as 'texts', portable linguistic objects. In the case of discourse analysts, this usually involves two discrete activities, one in which discourse is 'collected' with the aid of some kind of recording device, and the other in which the recording is transformed into some kind of artifact suitable for analysis.

Practices of entextualization have historically defined elite communities in society-- scribes, police officers, researchers-- who, through the 'authority' of their entextualizations are able to exercise power over others. To create texts is to define reality.

Whether we are talking about discourse analysts making transcripts or police officers issuing reports, entextualization normally involves at least six processes 1) framing, in which borders are drawn around the phenomenon in question, 2) selecting, in which particular features of the phenomenon are selected to represent the phenomenon, 3) summarizing, in which we determine the level of detail with which to represent these features, 4) resemiotizing, in which we translate the phenomena from one set of semiotic materialities into another, and 5) positioning, in which we claim and impute social identities based on how we have performed the first four processes.

These processes are themselves mediated through various 'technologies of entextualization' (Jones, 2009), tools like tape recorders, video cameras, transcription systems and computer programs, each with its own set of affordances and constraints as to what aspects of a phenomenon can be entextualized and what kinds of identities are implicated in this act. Changes in these technologies result in changes in the

practice of entextualization itself, what can be done with it, what kinds of authority adheres to it, and what kinds of identities are made possible by it.

Data in the audio age

The act of writing down what people say was pioneered as a research practice at the turn of the 20th century by anthropologists and linguists working to document the phonological and grammatical patterns of 'native' languages. Up until 50 years ago, however, what people actually said was treated quite casually by the majority of social scientists, mostly because they lacked the technology to conveniently and accurately record it. On the spot transcriptions and field notes composed after the fact failed to offer the degree of detail necessary to analyze the moment by moment unfolding of interaction. The 'technologies of entextualization' necessary to make what we now know as 'discourse analysis' possible were not yet available.

This all changed in the 1960s when tape recorders became portable enough to enable the recording of interactions in the field. According to Erickson (2004), the fist known instance of recording spoken interaction for research purposes was by Soskin and John in 1963 and involved a tape recorder with a battery the size of an automobile battery placed into a rowboat occupied by two arguing newlyweds. By the end of the decade, the problem of battery size had been solved and small portable audio recorders became ubiquitous, as did studies of what came to be known as 'naturally occurring talk', a class of data which, ironically, did not exist before tape recorders were invented to capture it (Speer 2002).

The development of portable audio-recording technology, along with the IBM Selectric typewriter, made the inception of fields like conversation analysis, interactional sociolinguistics, and discursive psychology possible by making accessible to scrutiny the very features of interaction that would become the analytical

objects of these fields. The transcription conventions analysts developed for these disciplines arose from what audio tapes *allowed* them to hear, and these affordances eventually became standardized as practices of 'professional hearing' (Ashmore et al. 2004) among these analysts.

The introduction of these new technologies of entextualization brought a host of new affordances and constraints to how phenomena could be framed, what features could be selected for analysis, how these features could be represented, the ways meanings could be translated across modes, and the kinds of positions analysts could take up *vis-à-vis* others.

Framing refers to the process through which a segment of interaction is selected for collection. Scollon and Scollon (2004) use the term 'circumferencing'. All data collection, they argue involves the analyst drawing a 'circumference' around phenomena, which, in effect, requires making a decision about the widest and narrowest 'timescales' upon which the interaction depends. All interactions are parts of longer timescale activities (e.g. relationships, life histories), and are made up of shorter scale activities (e.g. , turns , thought units). The act of 'circumferencing' is one of determining which processes on which timescales are relevant.

Among the most important ways audio recording transformed the process of framing for discourse analysts was that it enabled, and in some respects compelled them to focus on processes occurring on shorter timescales at the expense of those occurring on longer ones. One reason for this was that tapes themselves had a finite duration, and another was that audio recordings permitted the analyst to attend to smaller and smaller units of talk.

This narrowing of the circumference of analysis had a similar effect on the processes of selecting and summarizing that went in to creating textual artifacts from

recordings. Selecting and summarizing have to do with how we choose to represent the portion of a phenomenon around which we have drawn our boundaries. Selecting is the process of choosing what to include in our representation, and summarizing is the process of representing what we have selected in greater or lesser detail.

The most obvious effect of audio recording technology on the processes of selecting and summarizing was that, since audiotape only captured the auditory channel of the interaction, that was the only one available to select. While many researchers accompanied their recordings with notes about non-verbal behavior, these notes could hardly compete with the richness, accuracy, and 'authority' of the recorded voice. As a result, speech came to be regarded as the 'text'—and all the other aspects of the interaction became the 'context'.

It is important to remember that this privileging of speech in the study of social interaction was largely a matter of contingency. Analysts privileged what they had access to. Sacks himself (1984: 26) admitted that the 'single virtue' of tape recordings is that they gave him something he could analyze. 'The tape-recorded materials constituted a "good enough" record of what had happened,' he wrote. 'Other things, to be sure, happened, but at least what was on the tape had happened.'

While limiting what could be selected, the technology of audio recording hardly simplified the selection process. Because tapes could be played over and over again and divided into smaller and smaller segments, the amount of detail about audible material that could be included in transcripts increased dramatically. Whereas most analysts based their decisions about what features of talk to include on specific theoretical projects -- conversation analysts, for example, focusing on features which they believed contributed to the construction of 'order' in talk -- some analysts, like DuBois (DuBois et al 1993) promoted the development of more exhaustive systems

of transcription which not only fit present analytical interests but anticipated future ones.

One thing for sure was that the dramatic increase in the detail that could be included in transcripts had the effect of making discourse analysis seem more 'scientific', and over the years, the amount of detail in an analysts' transcripts came to be seen as a criterion by which the 'accuracy' of their data and the 'objectivity' of their work was judged. As Mishler (1991:206) describes it,

> researchers (strove) for more precision, detail, and comprehensiveness—pauses to be counted (by proper instruments) in hundreds rather than tenths of a second, the inclusion of intonation contours-as if that would permit us (finally) to truly represent speech.

This desire to 'truly represent' speech was thoroughly grounded in positivist assumptions about reality – that there was something objectively occurring to represent – assumptions which would soon rub up against the more dialogic and constructionist theories that were arising from these same studies of talk-in-interaction (Scollon 2003). As transcripts revealed to analysts the contingent and negotiated nature of talk, analysts were themselves forced to confront the contingent and negotiated nature of their transcripts. More recently, analysts seem to be weighing in on the side of variety rather than standardization (Bucholtz 2007) and selectivity over comprehensiveness (Duranti 2005). Analysts like Jaffe (2007), in fact, have gone so far as to suggest that <u>less</u> delicate transcripts might in some cases constitute more 'accurate' representations of participants' 'voices'.

<u>Resemiotization</u> is the process through which we translate phenomena from one set of semiotic materialities into another (Iedema 2001). Meanings are expressed differently in different semiotic systems, and so they cannot simply be transferred

from one mode to another; they must be 'translated.' In data collection using audio recorders, for example, the social interaction, what is essentially a rich multimodal affair, is resemiotized into a mono-modal audiotape, later to be further resemiotized into a different mono-modal artifact, a written transcript. In this process, the spatial and temporal aspects of the dynamic, multimodal interaction must somehow be 'translated' into the static, linear, and mono-modal materiality of text.

One important aspect of resemiotization in written transcripts is how the spatial arrangement of the page acts to translate certain temporal and relational aspects of the original interaction. Although there have been a number of experiments in the written representation of interaction using non-standard layouts and notations (see for example Ochs 1979, Erickson 2003), most transcription systems developed for audio data are arranged in the conventional 'play-script' layout, a layout that has a number of important effects on how we experience the interaction. First of all, the format creates the impression that interaction is focused, linear and monofocal, masking any simultaneity of action, nonlinearity, or polyfocality that might have been part of the actual interaction. Second, it impies a contingent relationship between immediately adjacent utterances of different speakers, whether or not one actually exists (Ochs 1979). Finally, it imposes on the transcript a particular 'chronotrope' (Bakhtin, 1981) or felt 'time-space' that may be radically different from that of the original interaction. In fact, one of the most jarring discoveries of those coming fresh to discourse analysis is how much longer it takes to read through the transcript of an exchange, with all of its details arranged linearly down the page, than it took the participants to actually produce the exchange. In short, the 'play-script' format requires that the reader rely primarily on the narrative interpretation of the analyst embodied in the sequential emplacement elements on the page to make sense of what

happened.

Perhaps the most important process of entextualization, at least that with the most obvious social consequences, is positioning. Whenever we turn phenomena into a text, we are making claims as to who we are and what our relationship is to those whose words and actions we entextualize and those with whom we will later share these entextualizations.

One rather obvious way that practices of data collection and transcription position the analyst is in how they reveal his or her affiliation to a particular 'school' of discourse analysis. It is, in fact, possible to give a cursory glance to a transcript and predict the kinds of theoretical positions about language the analyst will be advancing. As Jaffe (2007) has pointed out, transcription has become a kind of 'literacy practice', the mastery of which has become necessary for admittance into certain communities of scholars.

Beyond signaling disciplinary affiliation, however, the new forms of transcription that audio recording made possible to discourse analysts also made possible for them new positions of authority viz-a-viz their various audiences such as colleagues, tenure boards and funding bodies, as well as their 'subjects'. This authority came, first, from the level of detail they were able to present in their transcripts, which they could use an emblem of 'expertise'. Bucholtz (2000) has shown how the use of special fonts and annotations work to 'technologize' a text, and in the process, confer an identity of scientific expertise on the author.

This new authority also came from the 'evidentiary' nature of the tape itself as a material object, the notion that by possessing the recording the discourse analyst had access to 'what really happened' against which both the 'authenticity' of the transcript and any claims or counter-claims about it could be measured. Ashmore and his

colleagues (2004) call the tendency to confer on 'the tape' an epistemic authority 'tape fetishism'. The most dangerous thing about such an attitude is not just that the supposed 'authority' and 'objectivity' of 'the tape', produced as it was in particular circumstances of recording and listened to in varying contexts of hearing, are so easily called into question, but also because the existence of the recording itself lends further authority to the transcript, which is presumed to be the 'child' of the tape. This overconfidence in recording and and transcripts in the domain of discourse analysis simply makes for sloppy work. In other domains like law enforcement (Bucholtz 2009) the consequences can be rather more serious.

With the new authority granted to discourse analysis by the invention of the portable tape recorder, there also came new responsibilities. For one thing, analysts found themselves embedded in a complex new set of ethical and legal relationships with the subjects of their analysis. Much of the pioneering work using audio recorders simply ignored this complexity—it is hard to imagine, for example, how Sacks's recording of suicide hotlines would be treated in light of today's standards of 'informed consent'. Eventually, however, ethics boards and the law caught up with us. Not only do institutional review boards now demand that informed consent be obtained from any party whose voice is tape-recorded, but in many countries the law also demands it. These constraints, have left discourse analyst struggling to find ways to preserve the 'naturalness' of interactions in which all involved are aware they are being recorded, a most 'unnatural' state of affairs. The great irony of recording technology for discourse analysts is that it simultaneously introduced a standard of 'naturalness' for our data and created social and institutional conditions that made that standard much more difficult to obtain.

Video killed the discourse analyst?

Audio recording was not the only technology social scientists used in the mid-twentieth century to study communication. As early as the 1940s, Gregory Bateson and Margaret Mead (1942) were pioneering the use film in the study of communication, a technique that was later adopted by Edward Hall (1963) in his early studies of proxemics. By the 1970s, analysts like Birdwistell (1970) had begun to develop transcription systems for non-verbal features of social interaction. The assumption of these analysts was that meaningful interaction proceeds not just through talk, but through a host of other behaviors as well. This assumption would nowadays be considered non-controversial, but, in the 1960s and 70s, it failed to gain much traction, not until, of course, the invention of the video camera, a new 'technology of entextualization' capable of capturing not just words but also bodies in motion in a much cheaper and more immediate way than earlier film technology. Discourse analysis was ruined forever.

Only in one sense, that is. Discourse analysts could no longer just pay attention to phenomena that had traditionally been labeled 'discourse'; they could no longer ignore non-verbal behavior, which played so demonstrably an important role in all social interactions. And the technology that allowed analysts access to that behavior involved a whole new set of processes through which discourse analysts could frame, select, summarize, and resemiotize their data and position themselves in relation to it.

One important change came with the analyst now being able to frame his or her data spatially as well as temporally. With audiotape, only the duration, the starting point and ending point of the interaction mattered. Now the interaction had to be framed in space as well, with a whole new set choices to be made about what and whom should be included in the frame, the angle at which it should be shot, and so

forth.

Video also made the choices involved in selecting and summarizing much more complex, as nearly every aspect of non-verbal communication from gesture to gaze to body movement could be considered potentially communicative, as could a whole host of other non-verbal cues like dress and built environment. The biggest difficulty, however, came in the process of resemiotization, the challenge of translating the rich, multidimensional display of videotape to the still-dominant two dimensional medium of the written transcript (Park and Bucholtz 2009).

Early users of video essentially treated it as an extension of the audio recorder, using it as an aid to adding information about such things as gesture and gaze as notations within what were essentially conventional audio transcriptions (see for example Goodwin, 1986, Ochs and Taylor 1992). Many early attempts at multimodal transcription, were hindered by the essentially 'verbal logic' of the 'play script' model which analysts had inherited from the audio days, a model which provided few resources for representing the complex timing and simultaneity of actions and words in multimodal interaction. The problem with most early work using video was that technologies of transcription had not yet caught up with technologies of recording.

At the same time, video introduced further complexity into the analyst's relationships with other people. Since video data so clearly identify their objects, it became much more difficult to promise anonymity and confidentiality to participants. Furthermore, the 'gaze' of the camera in many ways turned out to be much more intrusive than the 'ear' of the tape recorder, giving rise to new layers of self consciousness and artificiality compromising the 'naturalness' of our data. Video technology also had an effect on the analyst's relationship with the consumers of his or her data, particularly the publishers of academic journals and books who were in

those early years still reluctant to incur the extra expense of publishing the photographs and other visual data many analysts found essential for communicating their findings, and the print medium itself, still the only medium that seemed to garner any recognition from academic institutions, lacked the ability to give readers access to anything but static images.

In the 1980s and 90s, the constraints and complications of video recording often seemed to outweigh the dramatic new affordances the medium offered, and many analysts, despite overwhelming evidence of the importance of the visual channel in social interaction, held stubbornly to the mono-modal talk-based approach to interaction which had served them so well in the past. This, however, was soon to change.

Data collection and transcription in the digital age

Many of the issues that plagued early users of video began to be resolved at the turn of the century as analysts like Baldry and Thibault (2006) and Norris (2004) began devising fully theorized systems of multimodal transcription. These breakthroughs, however, were not all theoretical. They came as well from another dramatic material change in the 'technologies of entextualization' available to the analyst, a change that was made possible by the digital revolution.

The qualitative difference between analog recording and digital recording as technologies of entextualization cannot be overstated. First of all, as digital video cameras shrunk in size, and as the practice of shooting digital video became more and more ubiquitous in the general population, the inconvenience and 'weirdness' of collecting video data decreased considerably. In addition, digital recording tended to deliver much higher quality outputs than earlier analogue systems, and increases in the size of computer drives and other solutions such as such as cloud storage helped

researchers to overcome difficulties in storing and backing-up their data.

Changes in social practices around video recording, not just among analysts, but among participants themselves, also introduced a range of new possibilities around the framing and selection of data. In particular, opportunities arose to record interactions not just from the point of view of the researcher but from the point of view of participants using micro-portable wearable cameras (Chalfen, 2014) or to engage participants in gathering data themselves using their mobile phones. Such techniques allowed researchers to more easily capture the mobile dimension of many social interactions (McIlvenny 2014, Mondada, 2014) as well as giving them access to more emic, experiential and embodied perspectives. Meanwhile, interactions recorded by people outside of research contexts and shared on social media sites such as YouTube became a new source of 'naturally occurring' data for discourse analysts (Jones, 2016).

Finally, digital media have dramatically altered the materiality of social interaction itself. One the one hand, many digitally mediated interactions (such as chats, instant messages, and interactions on social media sites) are already produced though written text, allowing researchers to sidestep the challenges of transcription altogether, but introducing new challenges around the selection and recontextualization of data as well as ethical associated with collecting it. On the other hand, new modes of mediated embodied interaction using video chat technologies (such as FaceTime and Zoom) introduce new theoretical challenges for multimodal interaction analysts in accounting for the different ways people manage things like gaze, turn-taking and the use of physical space online.

Among the most important affordances of digital audio and video recording for the discourse analyst is that it can be handled and manipulated in so many

different ways, many of which are reminiscent of the ways we handle and manipulate written text—it can be searched, tagged, annotated, chopped up, rearranged, and mixed with other texts. Because of this, it creates a host of new affordances when it comes to transcription. In other words, digital video has not just changed how analysts are able to record video, but also what they are able to do with it afterwards, how they are able to transform recordings into objects of 'analytic utility'.

The ability to easily capture still images from video meant that analysts no longer had to rely solely on text to describe behavior. Text and images could be integrated in ways that made transcripts themselves 'multimodal'. The practice of including still images captured from digital video in transcripts has been developed to great sophistication by scholars like Baldry and Tibault (2006) and Norris (2004).

Such 'multimodal transcripts', however, are still not the most multimodal means we have at our disposal to represent our data. Gu (2006), for example, has promoted the use of a 'corpus friendly' digital multimedia system for representing interaction which avoids the need for orthographic transcription altogether, and software solutions like Transana and Elan allow analysts to integrate their videos with their transcripts, their coding and their notations in flexible, searchable ways (Mondada 2009). Such advances have given to analysts the feeling that they are closer to the 'reality' of the original interaction than ever before.

But they are not. They are closer to a digital fabrication of reality, which is still only a fabrication. Just as 'tape fetishism' led analysts in the audio age to believe that 'accurate' and exhaustive transcripts of tape recordings would allow them to once and for all *truly* represent speech, the ability these new digital solutions give to analysts today to analyze video 'directly', seemingly unmediated by the transcription process, creates the illusion that they do not have to truly represent anything, (that the

video has done that for them), the illusion that the complex problems of selecting and summarizing can be somehow side-stepped, and that the inevitable distortions that accompanied the transformation of audio tape to written text can now be completely avoided, in short, the illusion that the age of 'transcriptionless analysis' has arrived.

As Mondada (2009) has pointed out, however, the viewing, coding and manipulating of video data with such software packages is far from unmediated. Users still need to go through the same five processes of entextualization that transcribers apply to audiotape. They still need to determine what counts for them as a meaningful unit of social interaction; aspects of the data still need to be selected, coded or otherwise summarized; and videos are still resemiotized into complex 'semiotic aggregates' combining symbols and writing with the audio and visual modes of the video. Unlike written transcripts, multimodal texts have no readymade 'textual units' apart from time codes, and so analysts must invent new ways to divide up the dynamic stream of behavior into manageable, intelligible bits. And these products of entextualization need to be still further entextualized into objects that can be published in an academic press still dominated by the medium of print.

Software can impose just as sturdy a set of theoretical assumptions on the analyst as a transcription system. All entextualizations are necessarily arrived at dialogically and are thus inherently 'double-voiced' (Bakhtin 1984: 185). By loosing sight of this 'double-voicedness', by thinking they can sidestep the gap between the original and the entextualized, discourse analyst are in danger of regarding their 'multimodal transcripts' as somehow more objective and transparent.

The notion that a 'multimodal transcription' of a video is necessarily a more 'accurate' portrayal of 'reality' than a careful transcription of an audiotape is really a matter of opinion. It depends primarily on how one defines 'reality'. Both annotated

video and written transcripts are artifacts, products of complex processes of framing, selection, summarizing and resemiotization, whose meanings change as they are transported across boundaries of time, space, and media (Jaffe 2007).

## Conclusion

In his article 'The Dialogist in a Positivist World', Scollon (2003) explores the balancing act discourse analysts have to perform to avoid, on the one hand, over-reifying their data and falling into a naive positivism, and, on the other hand over-relativizing their data and sinking into deconstructive impotence. The chief concern for a discourse analyst, he argues, is how to 'produce a working ontology and epistemology that will underpin (his or her) wish to undertake social action' without buying into the social constructions that underpin this action (p. 71). The sometimes-paradoxical history of data collection and transcription in discourse analysis is really the history of this dilemma.

The lesson of this history for anyone starting out in discourse analysis is that no technology of entextualization can capture the universe (Cook 1990). Nor is this what we need. The whole reason for entextualization is not to reproduce the universe, but to re-present it, and, by doing so, to understand it better. And it is these very processes of framing, selecting, summarizing, resemiotizing and positioning that allow us to arrive at these understandings.

Too often analysts have taken a 'deficit' attitude towards entextualization, lamenting how much of the 'original' interaction was 'lost in the transcription'. The fact is, what we search for in our transcripts is not 'truth', but rather 'analytic utility', Their ability to help us answer the questions we have about human communication and social interaction, not the degree to which they 'resemble reality', should be the

main criterion for judging the value of our transcripts.

At the same time, we must never loose sight of the ways technologies of entextualization profoundly affect our relationships with those whose words and behavior we study. The better our technology has become at capturing the details of social interaction, the more pressing and complex have become the ethical issues surrounding the activities of data collection and transcription. As Scollon and Levine (2004:5) write:

> The primary question now is not: Do we have or can we develop the technology needed to record the behavior of others? The primary question is: What rights does an academic researcher have in relationship to and in negotiation with her or his subjects of study? … In short, can our data collection and our analyses do others good or harm, and can we control those outcomes?

Key readings


1. Edwards, J.A. and Lampert M. D. (eds.) (2014), Talking Data: Transcription and Coding in Discourse Research 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.
2. Jones, R.H. (2016) Spoken Discourse. London: Bloomsbury.
3. Lapadat, J.C. and Lindsay, A.C. (1999), 'Transcription in research and practice: from standardization of technique to interpretive positionings', Qualitative Inquiry, 5, 64-86.
4. Norris, S. (2002). 'Implications of visual research for discourse analysis: Transcription beyond language', Visual Communication, 1, 97-121.
5. Ochs, E. (1979), 'Transcription as theory', in E. Ochs and B. Schieffelin

(eds), Developmental Pragmatics. New York: Academic Press, pp. 43-72.

References

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

Ashmore, M. and Reed, D. (2000), 'Innocence and nostalgia in Conversation Analysis: the dynamic relations of tape and transcript'. In: *Forum: Qualitative Social Research 1*(3). Available at < http://www.qualitative-research.net/index.php/fqs/article/view/1020/2199>, accessed 8 January 2010.

Bakhtin, M. (1981),, The Dialogic Imagination. (C. Emerson, trans.). Austin: University of Texas Press.

Bakhtin, M. (1984), Problems of Dostoevsky's Poetics (C. Emerson, trans./ed.). Minneapolis: University of Minnesota Press.

Baldry, A. and Thibault, P.J. (2006), Multimodal Transcription and Text Analysis. London: Equinox.

Bateson, G. and Mead, M. (1942), Balinese Character, A Photographic Analysis. New York: The New York Academy of Sciences.

Bauman, R. and Briggs, C.L. (1990), 'Poetics and performance as critical perspectives on language and social life', Annual Review of Anthropology, 19, 59-88.

Birdwhistell, R. (1970), Kinesics and Context. Essays on Body Motion Communication. Philadelphia: University of Pennsylvania Press.

Bucholtz, M. (2007), 'Variation in transcription', Discourse Studies, 9, 784-808.

Bucholtz, M. (2009), 'Captured on tape: professional hearing and competing extextualizations in the criminal justice system', Text and Talk, 29, 503-523.

Chalfen, R. (2014) '"Your panopticon or mine?" Incorporating wearable technology's Glass and GoPro into visual social science',*Visual Studies*, 29, 299-310.

Cook, G. (1990), 'Transcribing infinity: Problems of context presentation', Journal of Pragmatics, 14, 1-29.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., and Paolino, D. (1993), 'Outline of discourse transcription', in J. A. Edwards and M. D. Lampert (eds), Talking Data: Transcription and Coding in Discourse Research. Hillsdale, NJ: Lawrence Erlbaum, pp. 45-87.

Duranti, A. (2006), 'Transcripts, like shadows on a wall', Mind, Culture and Activity, 13, 301-310.

Edwards, J. A. (1993), 'Principles and contrasting systems of discourse transcription', in J. A. Edwards and M. D. Lampert (eds), Talking Data: Transcription and Coding in Discourse Research. Hillsdale, NJ: Lawrence Erlbaum, pp. 3-31.

Erickson, F. (2003), 'Some notes on the musicality of speech', In D. Tannen and J. Alatis (eds), Linguistics, Language, and the Real World: Discourse and Beyond. Washington, D.C.: Georgetown University Press, pp. 11-35.

Erickson, F. (2004), Origins: A brief intellectual and technological history of the emergence of multimodal discourse analysis', In R. Scollon and P. Levine (eds), Discourse and Technology: Multimodal Discourse Analysis. Washington, D.C.: Georgetown University Press, pp. 196-207.

Goodwin, C. (1986), 'Gestures as a resource for the organization of mutual orientation', Semiotica, 62, 29–49.

Gu Y. G. (2006), Multimodal text analysis: A corpus linguistic approach to situated discourse. *Text and Talk*. 26, 127-167.

Hall, E.T. (1963), ‘A system for the notation of proxemic behavior’, American Anthropologist, 65, 1003–1026.

Iedma, R. (2001), ‘Resemiotization’, Semiotica, 137, 23-39.

Jaffe, A. (2007), ‘Variability in transcription and the complexities of representation, authority and voice’, Discourse Studies, 9, 831-837.

Jones R.H. (2009) ‘Dancing, skating and sex: Action and text in the digital age’, Journal of Applied Linguistics, 6, 283-302/

McIlvenny, P. 2014. “Vélomobile Formations-in-Action: Biking and Talking Together.” Space and Culture, 17, 137–156.

Mishler, E. G. (1991) ‘Representing discourse: The rhetoric of transcription’, Journal of Narrative and Life History, 1, 255-280.

Mondada, L. (2009), ‘Commentary: transcript variations and the indexicality of transcribing practices’, Discourse Studies 9, 809–821.

Mondada, L. (2014) ‘Bodies in action: Multimodal analysis of walking and talking’, Language and Dialogue, 4, 357-403.

Norris, S. (2004), Analyzing Multimodal Interaction: A Methodological Framework. London: Routledge.

Norris, S. and Jones, R. (eds.) (2005), Discourse in Action: Introducing Mediated Discourse Analysis. London: Routledge.

Ochs, E. and Taylor, C. (1992), ‘Family narrative as political activity’, Discourse and Society, 3: 301–40.

Park, J.S. and Bucholtz, M. (2009), ‘Introduction: Public transcripts: entrxtualization and linguistic representation in institutional contexts’, Text and Talk, 29, 485-502.

Psathas, G., and Anderson, T. (1990), ‘The "practices" of transcription in conversation analysis’, Semiotica, 78, 75-99.

Sacks, H. (1984), ‘Notes on methodology’, in J.M. Atkinson and J. Heritage (eds), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, pp. 21–7.

Scollon, R. (2003), ‘The dialogist in a positivist world: Theory in the social sciences and the humanities at the end of the twentieth century’, Social Semiotics, 13, 71-88.

Scollon, R. and Levine, P. (2004), ‘Multimodal discourse analysis as the confluence of discourse and technology’, in R. Scollon and P. Levine (eds), Discourse and Technology: Multimodal Discourse Analysis. Washington, D.C.: Georgetown University Press, pp. 1-6.

Scollon, R. and Scollon, S. W. (2004), Nexus Analysis: Discourse and the Emerging Internet. London: Routledge.

Soskin, W. F. and John, V. P. (1963), ‘The study of spontaneous talk’, In R. Barker (ed), The Stream of Behavior. New York: Irvington Publishers, pp. 228-281.

Speer, S. (2002), ‘Natural' and “contrived” data: A sustainable distinction?’, Discourse Studies, 4, 511-525.