Collection, handling, and analysis of classroom recordings data: using the original acoustic signal as the primary source of evidence

Montserrat Pérez-Parent

School of Linguistics & Applied Language Studies, The University of Reading

Abstract. The present paper describes the methods used for data collection, handling and analysis of data on the research project "Talk about texts in the guided work period of the Literacy Hour". This project was funded by the Economic and Social Research Council (award reference R000223563) and investigated the nature of talk between teachers and pupils during the guided group work period of the Literacy Hour in year 6. The paper describes how the research fulfilled one of its objectives, namely that of demonstrating a specific innovation in research methodology for corpus formation and computer-assisted analysis of naturalistic discourse. The innovative methodology described uses the original acoustic signal as the primary source of evidence, with supplementary transcripts and analytical coding mapped directly on to this signal. The paper also describes some of the quantitative analyses that the collection of data in the manner described allowed the research team to make.

1. Introduction

A number of frameworks have been devised for the analysis of classroom discourse (Sinclair and Coulthard 1975; Mehan 1979; and the survey in Edwards and Westgate 1994). Whilst some specific concepts from this tradition of research, such as the Initiation-Response-Evaluation (IRE) structure typical of teacher-led discussion, may have informed analysis in the present study, it was not possible to take over any of the existent frameworks intact, since these were produced in other situational contexts and with different theoretical purposes in mind. Rather, the present study seeked, and achieved, to develop a new analytical framework directed at interpreting the educational significance of moves in the discussion considered as evidence of pupils' powers of comprehension.

2. The recordings: settings, procedure and equipment

The primary data source for the study were audio recordings of teacher-pupil discussion in the guided reading session of the literacy hour. These

© 2002 M. Pérez-Parent

were collected at five different primary schools in the Swindon area. With the members of the research team being based at Reading and Bath, Swindon proved to be a convenient location which allowed the team members quick access to the schools by the time lessons started without involving excessive travelling. The identities of the schools have been kept anonymous in any reporting of the project results. A three-letter code was used which would only reflect the identity of the school to the researchers and the teacher and head-teacher of that particular school, while making it unrecognisable to others. The codes used for the five participating schools were: FND, WTL, STC, KWM and ABM.

A sample of five schools was considered to be sufficiently large to make the comparison between the different sites of theoretical interest, and to avoid the possibility of unusual school conditions distorting the findings. As it turned out, the data from one of the schools (STC) had to be discarded from the analysis due to teacher not adhering to the guidelines for the recording sessions that had been set up. The resulting recordings at this particular school were far much longer than required and the tasks carried out by the pupils were not spontaneous enough. Thus, they were not comparable with the data obtained at the other four schools.

Each school was visited on three occasions staggered over a six-month period interval (October 2001, January 2002 and March 2002). This provided a longitudinal element to the research design, which made it possible to explore evidence of the development of pupils' powers of comprehension over this period. Educational research literature also suggests that building up a shared history of collaboration is a factor in the effectiveness of pupils' joint working (Halligan 1988), and the series of visits enabled this factor to be taken into consideration. The successive visits to record the same group of pupils made them and the rest of the class become familiar with the setting up of the equipment and the recording procedure and both teacher and pupils gradually became more at ease with it as the visits went by.

The recordings, which lasted for 20-30 minutes on each occasion, were of the same group of 6 children in year 6 (10–11 year old pupils) and their teacher during the guided group work period of the Literacy Hour, with the exception of one recording session in which one of the pupils was absent due to illness (WTL05-03-02). The three members of the research team were present in the classroom during the recordings. Two[1] set up the equipment and oversaw the recordings, and the third one[2] assisted the rest

[1] Dr Simon Arnfield and Ms Montserrat Pérez-Parent (The University of Reading).

[2] Dr David Skidmore (The University of Bath).

of the class with their work while the teacher was conducting the group discussion. With the aim of minimising disruption to the class as much as possible, the team had designed a recording protocol that allowed setting up, and later on dismantling, all the equipment in about 5 minutes. The setting up of equipment took place either before the pupils entered the class or in between activities, for which a break was allowed and pupils had to move places around the classroom anyway. The goal of minimum class disruption was achieved successfully on every visit to each school. A pre-recoding protocol also ensured that the batteries on all equipment were charged and there were blank disks and tapes in the mini-discs and the camcorder. As a result, all recordings went smoothly and no data was lost. This protocol also established that the microphone and mini-disk attached to it corresponding to each speaker was colour coded. A record was kept of what colour each speaker got given so that they could be assigned the same colour in future visits. This allowed keeping track of which particular pupil produced what amount of data while keeping their anonymity.

Previous research has traditionally used a single tape recorder for this kind of recordings, with a multidirectional microphone placed in the middle of the table around which participants are sitting down. Because of the high levels of background noise going on in classrooms, using a single microphone meant that it was sometimes difficult to distinguish who was talking, let alone what they were saying. Also, overlaps and aside conversations between participants were lost. In order to achieve an accurate recording of everything that was going on within the group, the present project aimed at recording the teacher and each pupil individually. The use of a lapel microphone for each speaker allowed doing this while providing high resolution for speakers' utterances. Multiple simultaneous recordings would not have been possible with tape recorders, as they have been proved to record at different speeds and it would not have been possible to synchronise the multiple recordings later on. Mini-disc was chosen as the most appropriate medium for carrying out this kind of recordings, taking into account factors of quality, cost, durability and ease of use. Unfortunately, and contrary to expectations, mini-discs also turned out to record at different speeds and the resulting individual sound files had to be further processed and stretched to a common length before synchronizing them. To sum up, the individual recordings were thus made using a lapel microphone (unidirectional tie clip) and a portable mini-disc recorder (Sharp MD-MT90H(S)) for each participant.

Acoustic recordings alone, though, are not sufficient to document the discourse, as details such as gestures are not recorded. Traditionally paper based methods have been used to record these events which inevitably lead to omission and error. The recorder has to make real-time decisions as to what factors are relevant, and may miss events or may make mistakes. A video camera mounted on a tripod (Sony Handicam Vision Video Hi8 – CCD-TRV64E PAL) was used to record aspects of non-verbal communication during the discussion[3]. This is an improvement over paper records as the video data may be analysed at a later point or checked to verify points of contention within the data. It is also possible to make accurate timings and alignments between non-verbal events and the acoustic records.

3. Data handling: digitisation, processing and transcription

Each visit resulted in seven concurrent mini-disc recordings and one concurrent video recording. This data was digitised to provide computer readable format. The mini-disc recordings were digitised using a standard PC soundcard with the Cool Edit 2000 package (sample rate 32000, mono, 16bit) and stored as standard PCM Wav format files for backup, but were converted to MP3 format for final distribution. Each file's name consists of the three-letter code for the school, followed by the colour code of the speaker, and then the date the recording took place. Thus, file KWMgreen06-03-02.mp3 corresponds to the recording of the pupil who was assigned colour green at school KWM on the third visit made on 06 March 2002.

The video recordings (without sound) were digitised using a video capture card at a rate of 15fps 1/4 PAL image and encoded using a DivX codec. Both MP3 and DivX formats allow for small file sizes, which are efficient and well suited to world-wide-web base delivery, the method of final distribution of the corpus material. This consists of MP3 merged audio sample files, alongside the XML format label files.

At the commencement of each recording session a handclap was recorded which was picked up by all mini-discs and visible by the camcorder. This was to provide for synchronisation between the recordings; each recording was then trimmed to the maximal onset of the clap. The aligned audio samples were then merged to provide an overall audio track, which was then added to and synchronised with the video

[3] The video material was used to assist the research team in making interpretations of the discussion and to enhance annotations within the research project. Video samples have not, and will not, be made publicly available in consideration of ethical constraints and in particular to protect the identities of the subjects.

sample. Because of the variation of the tape speed of an analogue camcorder it was necessary to adjust the frame rate to ensure correct synchronisation between the audio and video signals.

An initial transcription of the individual audio samples (for each subject) was made by a trained transcriber. Because each subject was recorded individually, a high quality signal was recorded which allowed the accurate transcription of what every subject said free from constraints of the subject being overshadowed by other noises. This allowed the recording of everything that went on during the session, including pupils quietly talking to themselves, whispering, and conversations between pupils which run parallel to the main discussion being let by the teacher at that particular moment. Our method is innovative in the sense that it allowed for all these things to be recorded for the first time, which were not possible to pick up using a single multidirectional microphone.

The initial transcriptions were cross-checked and amended by the research team. They were then segmented into "utterances"[4] and aligned to the sample for that speaker using the Xwaves package. To reduce manual effort only utterance numbers were labelled using Xwaves and a program automatically modified the label files to insert the utterances from the transcription file.

In order to add annotations to the data that were observed from the video, it was necessary to convert the aligned labels into the format used by Anvil (Kipp 2001), the software package we used for this phase. Anvil will read Xwaves format label files but interprets them in a different way than desired, so the Anvil XML format files needed minor automatic modification to delete repeated elements. The use of Anvil allowed us to view the video, listen to the merged audio file and follow each individual transcription at the same time. It also allowed the creation and insertion of labels describing educationally relevant material.

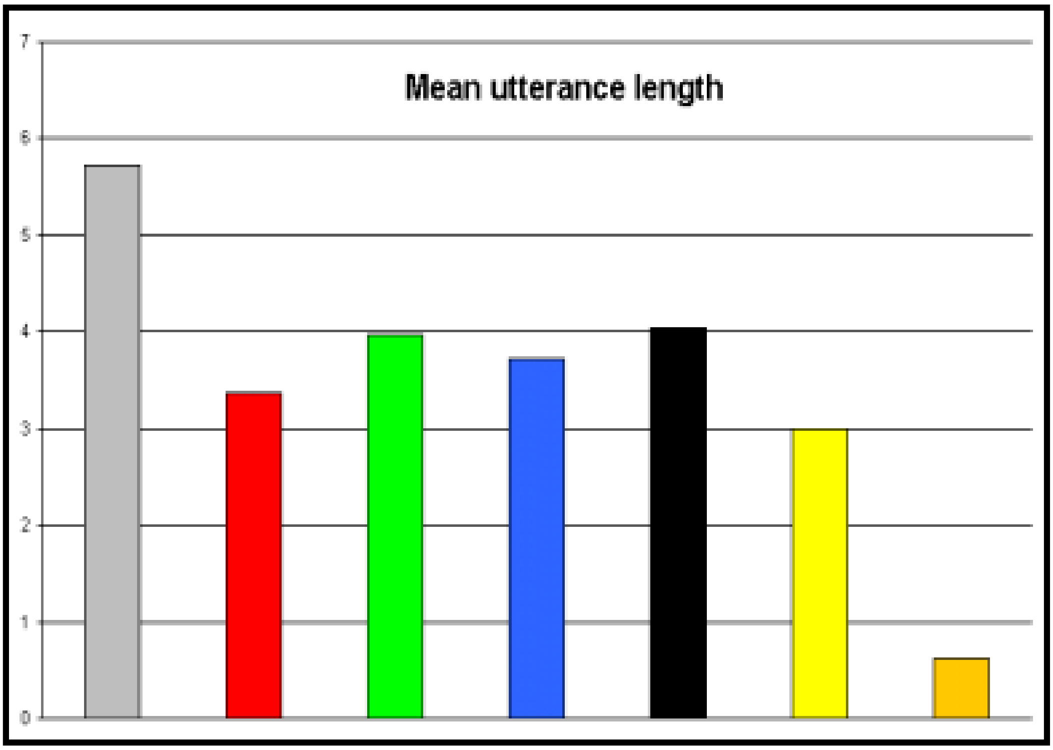
4. Quantitative analysis

Various statistics and graphs can be produced to characterise the data and give a visual overview of the dialogue as observed in our data. Below are some examples of the graphs that aided us in our analysis. In all the graphs, each pupil is depicted with the colour that they were assigned on

[4] Deciding what an utterance is is one of the biggest decisions that one has to make when transcribing and analysing spoken language. Although not without criticism, the operational definition of utterance that we followed in the present research is that of utterance as the speech continuum from one silence to the next.

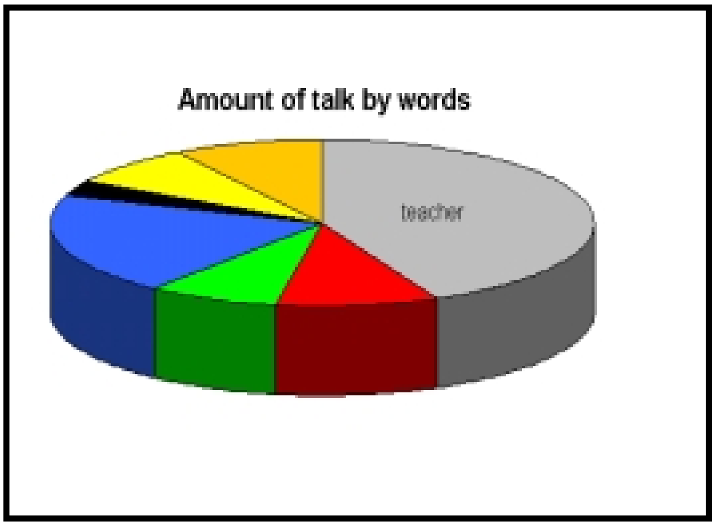
the recording session. The teacher’s colour is grey. All graphs correspond to the same data, namely FND23-10-01.

- Mean utterance length



We can see from this graph how the teacher’s utterances are longer that those of the pupils’. Most pupils average between 3 and 4 words per utterance, while the teacher constructs utterances with an average of almost 6 words. One of the pupils, gold, clearly shows a pattern contrasting with the rest of his/her classmates. The fact that gold shows such a low average of words per utterance means that s/he probably just produced one-word answers and backchannelling sounds to show the teacher s/he was following her explanations.

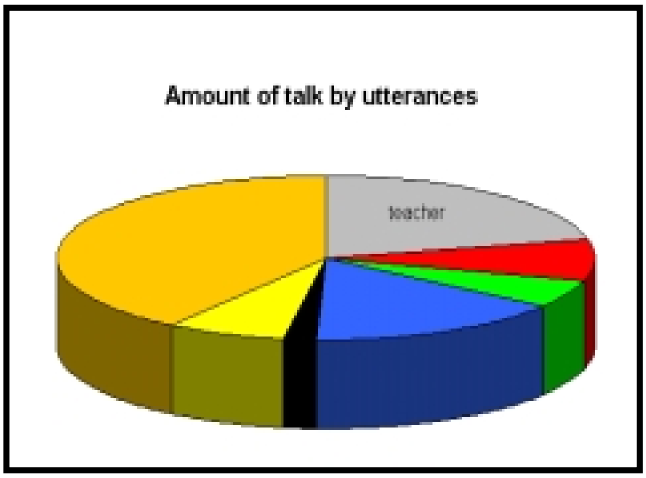
- Amount of talk by words



This pie chart shows the relative amount of words produced by each speaker and can show who actually does the most talking. As it could be expected, the figure shows that grey (the teacher) produces almost half of the speech in the conversation whereas the other speakers are all quite similar in the amount of talk they produce, with the exception of black who appears not to say as much.

It is interesting to see how gold, who produced very short utterances and thus presented the lowest score in the previous graph, here produces higher results than other pupils such as black or yellow. This means that, this particular pupil produced lots of speech but his/her contributions were mostly one-word answers rather than longer elaborations.

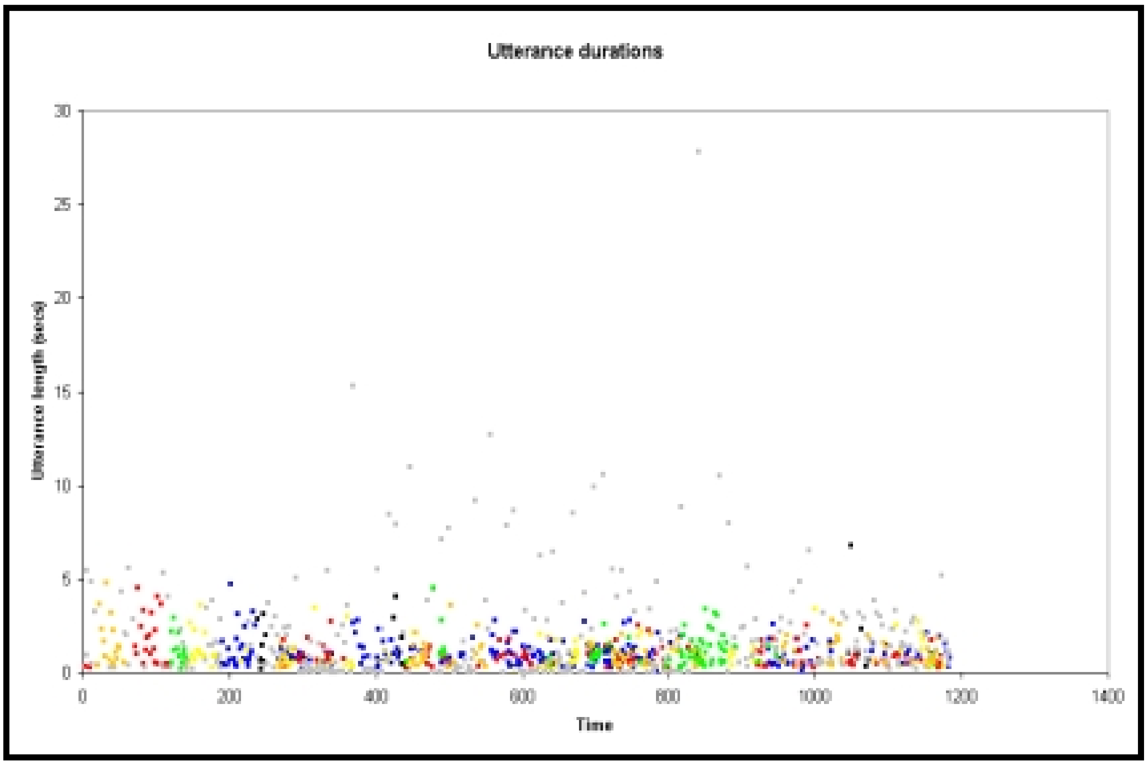
- Amount of talk by utterances



This pie chart shows the relative amount of utterances produced by each speaker and gives an indication of the amount of participation of each speaker in the conversation as opposed to the amount of speech actually produced. Whereas the previous figure indicates that the teacher does most of the talking, this graph shows that they do not actually speak significantly more times than the other speakers; it is just that the teacher's interventions are longer and more elaborated, thus consisting of more words.

Some speakers, for example gold, may produce many very short utterances, such as "yeah", "ok", "right", in response to what others are saying and this is shown by the fact that they produce many more utterances despite the previous graph showing that they say about as much as the other pupils. Black does not produce that many utterances, or words as shown by the previous graph, and hence they most likely only produce a few short utterances.

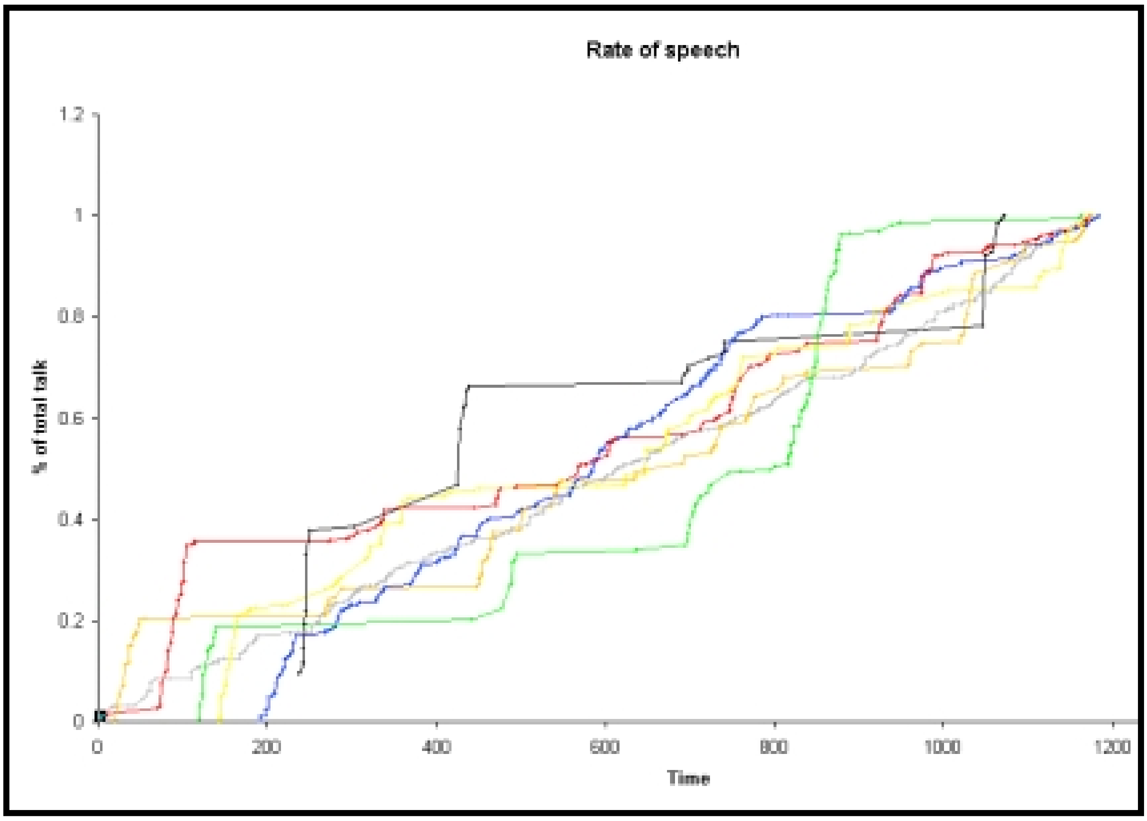
- Utterance durations



This figure shows another way of representing the duration of utterances, but this time along a time continuum. The Y-axis represents the duration of each utterance (in seconds), whilst the X-axis represents the position in time at which the utterance started. We can see how, again, the teacher's utterances are longer, as the grey dots appear higher on the Y-axis.

This graph also shows who is talking at which points throughout the conversation, as well as giving an indication of turn taking. Looking at the start of the conversation (between 0 and 200 seconds on the graph), it can be seen that each speaker talks in turn – grey, gold, red, green, yellow, blue and black – with interjections from grey (the teacher); the conversation is structured. As the conversation progresses, and especially between minutes 10 and 13 (between 600 and 800th seconds on the graph), we see how the interaction is not so clearly structured and the graph shows interventions from all speakers; conversation is more "free-form". At the end of the recording there is a clear dialogue between grey and blue.

- Rate of speech



This last figure shows the rate at which each speaker proceeds through their contributions to the conversation. The X-axis represents the timeline of the conversation. The Y-Axis represents the percentage of that speaker's speech (in terms of number of words). The colours represent each speaker, as in all the previous graphs. A point on a line shows the percentage of that speaker's total contribution to the conversation so far produced at the point in time as represented on the X-axis. When a speaker is not talking the line will be horizontal, and will rise (ultimately towards 100%) whenever the speaker talks.

References

Edwards, A. D., and D. P. G. Westgate (1994). *Investigating classroom talk*. (2nd edition). Lewes: Falmer.

Halligan, D. (1988). Is there a task in this class? In M. MacLure, T. Phillips and A. Wilkinson (eds.) *Oracy Matters*. Milton Keynes: Open University Press.

Kipp, M. (2001). Anvil - A generic annotation tool for multimodal dialogue. *Proceedings of Eurospeech 2001*. Aalborg. 1367-1370.

Mehan, H. (1979). *Learning lessons*. Cambridge, Mass.: Harvard University Press.

Sinclair, J. M., and R. M. Coulthard (1975). *Towards an analysis of discourse: the English used by teachers and pupils*. London: Oxford University Press.