Developing a Quality Spoken Component of the Australian National Corpus

Jean Mulder, Cara Penry Williams, and Deborah Loakes
The University of Melbourne

1. Introduction[1]

The creation of a quality spoken component of the Australian National Corpus (AusNC) will allow us to deepen our understandings of Australian English (AusE) and to open up new areas of analysis. To make the most of this opportunity we contend that not only must the data be of high quality but that the corpus must also be constructed in such a way that the data is of maximal use to researchers with differing interests.

Toward these aims, we begin by drawing on other spoken corpora and our collective experience to lay out a set of design principles for creating a quality spoken corpus (Section 2). Using a sample recording, we then demonstrate the value of the information which can be obtained through putting such design principles in place (Section 3). We conclude with what we believe is a viable proposal for generating a quality spoken component of the AusNC (Section 4).

2. Design Principles

Designing a quality spoken corpus that meets the needs of as many diverse researchers as possible involves first assessing the types of information that such a corpus should provide and then figuring out how to frame a corpus so that it can do this. Here we consider the issues of quality of recordings, range of recordings, supporting documentation, transcription, and searchability of the recordings, supporting documentation and transcription. As an argument for best practice processes, detailed proposals in each of these areas are presented.

2.1. Recording Requirements

A downfall with many recordings made for linguistic analysis is their quality. While spectrographic analysis is not a priority for most projects without a phonetic and phonological scope, we argue that aiming for the best quality recordings possible is always beneficial. This is for two main reasons. First, unforeseen problems commonly arise outside the immediate aims of the particular project for which the recordings were made, and these can often be answered through close phonetic analysis. Second, projects may well be carried out by different researchers at a later stage on the data recorded for a vastly different purpose, so having the best quality recordings possible means that no type of analysis will be precluded.

There are in fact numerous publicly available documents from sound engineers, phoneticians, and other relevant sound professionals which discuss the merits and disadvantages of various types of equipment for audio recording of speech. Some particularly useful information is available in Ladefoged's (2003) book *Phonetic Data Analysis*, which was written primarily as a fieldwork companion but is useful for all linguists making audio recordings.

[1] The authors would like to thank Janet Fletcher, Saya Ike, Michael Haugh and the participants of the Designing the AusNC Workshop held at UNSW, Sydney, 4-5 December 2008 for valuable discussion of the ideas in this paper. None of them is responsible for what we have done with their advice.

The AusNC will need to make decisions regarding audio recordings based on what is required from the corpus. However, there are some minimum requirements that can be adhered to. Some particularly useful information from Ladefoged (2003, p. 17) is provided below:

> Comparing the different systems for making recordings leads to no easy conclusions. There are four properties that one wants from a good recording system…: (1) a good frequency response…; (2) a good signal / noise ratio…; (3) reliability and user-friendliness; and (4) the possibility of using the recordings for a long time.

Guidelines on the homepage of University College London Department of Phonetics and Linguistics (2009) describe the necessity (and ease) for recording high quality data in the current age, where technology is relatively cheap and easily accessible:

> …the audio recording equipment available today is of such good specification that it makes sense to try to ensure that all speech recordings are of sufficient quality for instrumental analysis. This means taking care about the choice of microphone, the recording environment, and the recording settings as well as the choice of recorder.

Taking these recommendations into account, we posit some preferences for recording equipment. These are general baseline preferences, and of course other project-specific factors would need to be considered:

- Recording devices need to be mobile. In particular, they should be both small and portable so that they are easily transportable.
- The output needs to be digital and easily transferred onto computers for analysis. Data that are digitized when recorded do not need to be transformed, which could cause loss of quality.
- Data need to be recorded so that frequencies of <u>at least</u> 11 kHz can be captured (and with no limit on the lower frequencies, i.e. a range of 0-11kHz). Higher frequencies should be recorded where possible. Many recording programmes in computers, as well as digital audio tape (DAT) recorders, allow choice of frequency range (e.g., Ladefoged, 2003, p. 17-18).

Aside from the recording device itself, choice of microphone is also important. There are many possible types that can be used, for example lapel, headset or table; uni- or omni-directional; dynamic, electret or condenser. The exact type preferred for use is, again, dependent on the needs of the project. However, a preference is for a microphone that connects easily to the recording equipment so that there is no need for an external power supply, and therefore ease of portability is maintained. A relatively new type of microphone recommended for high quality recordings is a USB microphone (see University College London, Dept. of Speech, Hearing & Phonetic Sciences, 2009), which is easy to use, can be directly connected to a computer, and omits the need for an audio interface (so maintains ease of portability).

To summarize, we have discussed various types of factors that need to be considered for corpus recordings. While we have highlighted the variability inherent in equipment types, we have also provided some minimum requirements that, aside from consideration for the AusNC corpus, should also be of use for students of linguistics planning to make audio recordings.

2.2. Range of Recordings

As we have asserted, a central aim is to design a corpus that is as useful to as many researchers as possible. To this end, it would be an advantage to have a collection of recordings occurring in diverse settings.[2]

[2] Also, due to the detailed attention we are suggesting be given to each recording from this collection, only those of the highest quality need be selected for transcription (a proposal for this selection process is outlined in Section 4).

Collecting spoken data from a range of contexts would mean interactions in particular contexts could be studied, making the corpus of interest to a greater number of researchers. The best way to achieve this is through multiple recordings of the same speakers, which would also allow for the exploration of intraspeaker variation. This intraspeaker variation may relate to individual speaker conditions such as whether the speaker was psychologically stressed or ill at the time of the recording, or the speaker was particularly emotive or tired. Intraspeaker variation may also be caused by speaking style, for example whether the speaker was completing a reading task, or engaged in spontaneous speech, and who the other participant(s) are and how they are related to the speaker. Difference in speaker performance can also be caused by recording conditions and the context the speech is recorded in; for example, a recording of a casual telephone conversation would be vastly different, both acoustically and linguistically, to a sociolinguistic interview conducted in a recording studio. Recordings of the same speaker in different contexts would be illuminating on issues such as style and register, and would also allow comparison of the acoustic effects of varying recording conditions.

While having a wide range of recordings of speakers in natural settings is optimal for various reasons, one area which could then become compromised is the quality of the recording. Some ways of maintaining a high recording standard are to ensure that the highest signal to noise ratio possible is being used (to minimize the amount of background noise recorded), and to be cautious where the microphone is placed in relation to the focus speaker (to ensure that speaker's voice is the captured as well as possible). Again, Ladefoged's recommendations (2003, p. 21-23) for fieldwork conditions can be taken into account when deciding on the optimal conditions for each setting.

In sum, data should always be of the maximum quality possible, not only in terms of the sound quality as affirmed in the previous section, but also in terms of representing a range of settings, including less self conscious interactions where naturalness is desired. As we discuss in the next section, data of maximum quality are also accompanied by a range of supporting documentation, including information about the participants and the context of the recording.

2.3. Supporting Documentation

We believe that the provision of in-depth supporting documentation will contribute to the continuing usefulness of the corpus. The supporting documentation includes speaker permissions, technical specifications of recordings (e.g., recording sample rate and resolution, and microphone used and placement), contextual information and biographical, social and regional information about the participants.

To exploit the full research potential of the recordings, information about context should be gathered and made easy to access. Knowing where the interaction took place, the location or seating plan of the speakers during the interaction, what the relationships between speakers are, who else was present and how often these sorts of interactions occur between these participants would enrich the data and their analysis. Such information is important for modern sociolinguistic approaches and more generally provides details of interest to qualitative researchers from a range of disciplinary backgrounds. The Wellington corpus of Spoken New Zealand English, for example, collected such information (Holmes, Vine, & Johnson, 1998).

The speaker information should be as detailed as possible without placing a great burden on participants and the person entering this information into the database and, of course, without risking making the speakers identifiable. Social and regional information are crucial because corpus data provide such an excellent opportunity to bring a diverse group of speakers into one dataset. Having personal information collected on recordings may allow new insights in research on variation within AusE.

Some previous corpora have neglected to collect this sort of information. For example, the *Dimensions of Australian English* project carried out by Monash University in the late 1990s only recorded the sex, birthplace, age, languages spoken at home and ethnic background of family members. Detailed information about individual speakers would be an advantage in that it would assist the exploration of a greater range of research questions. For instance, researchers interested in multilingualism may need to know the range of languages a speaker uses, the languages spoken by the speaker's parents, etc. Research looking at the usage of a particular feature may rely on knowing

where a speaker lives and has lived, their education level and which education systems they have been involved in. Additionally, forensic phonetic research, which focuses on variation within and between individual speakers, is dependent on such information. While this may sound onerous, we note that the Wellington Corpus of Spoken New Zealand English collects much of this information using fewer than 20 questions (Holmes et al., 1998).[3]

2.4. Transcription

The transcript is central to research on spoken language. As Edwards (2001, p. 321) makes the case:

> Recordings are essential tools in discourse research, but are not sufficient by themselves for the systematic examination of interaction. It is simply impossible to hold in mind the transient, highly multidimensional, and often overlapping events of an interaction as they unfold in real time.
>
> For this reason, transcripts are invaluable. They provide a distillation of the fleeting events of an interaction, frozen in time, freed from extraneous detail, and expressed in categories of interest to the researcher.

Along with recognising the critical roles of both the recording and the transcript of it, we must also acknowledge the importance of maintaining the link between the two. Minimally, this can be done by time codes in the transcript, but, preferably, the spoken component of the AusNC would be designed so that a researcher would be able to listen to any prosodic unit or turn by simply selecting that particular portion of the transcript.

Turning to the nature of the transcript, we contend that the goal is to create a written representation of the recorded speech event that makes it accessible to a wide range of researchers. As how and what we transcribe shapes what we see (Ochs, 1979; Du Bois, 1991), the transcription system used in the AusNC must be to a high enough level to readily allow a wide range of linguistic research, including discourse analytic research, as well as comparison with other national spoken corpora. While this latter point, in particular, involves issues of coding (also known as 'annotation' or 'tagging') and mark-up, our focus here is on the level of representation of spoken language in a written medium.

Drawing from the literature on discourse transcription design principles (see for example the papers in Edwards & Lampert, 1993), what we envision is a broad transcription system that encodes:

- Basic units in which utterances are articulated (prosodic units[4] and turns)
- Temporal sequencing of utterances (in succession and simultaneous)
- Intonation contours and their function (e.g., final, continuing, appeal)
- Accent and fluctuations in timing (e.g., tempo, pause and lengthening)
- Nonverbal noises (e.g., laughter, inhalation and throat clearing)
- Voice quality (e.g., whisper, voice of another and creaky)

In particular, it should be noted that such an approach to transcription recognizes prosodically defined 'chunks' (known as IUs) as basic units of speech.

It is further envisioned that such a transcription system would be 'foundational' in that researchers would be able to add further transcriptional detail later as specific to project needs.

Finally, the layout and choice of symbols, along with consistent encoding, would make it possible to easily convert between different formats allowing for flexibility of representation.

---

[3] Holmes et. al.(1998) also discuss the information which the form did not capture and they later realized would have been helpful.

[4] For ease of reference, following Chafe (1994) and Du Bois, Schuetze-Coburn, Cumming, and Paolino (1993), we call these prosodic units Intonation Units (IUs), fully recognizing that prosody involves much more than 'intonation'.

To illustrate that transcriptions of this kind are useful in yielding insights into, for example, the perceptual cues that participants rely on in conversational management, consider the following examples from the ART corpus of Australian talkback radio.[5] In (1), we have kept the transcription system used in the corpus.

(1) Carpet Python (ART ABCe4 [Caller 5])

ABCe4:[P1] Good good now tell us about this carpet python.
ABCe4:[C5] Yeah well Ric last week I think you were talking about the colourations on carpet pythons and uh I was coming down a hill on a my mountain bike one day and um this is a couple of years ago out behind Townsville in the Mt Stewart Road and I came across a I presume it was a carpet python but it was bright yellow and black. Really distinct markings and I I've never seen one in a book like that and I've I've attended snake identification courses I've never seen anything quite like it but <P1 you were talking about> last week.
ABCe4:[P1] Yeah dur he was talking about um you were talking about bright yellow ones last week Ric.

Such a transcription system makes it difficult, for instance, to see the structure of the interaction, especially how the overlapping speech is resolved as the beginning and ends of the overlapping speech are unmarked for the second speaker. In particular, this makes it impossible to discern whether or not the *but* in the last line of C5's turn is a Final *but* (see Section 3.3).

In (2), another example from the corpus has been re-transcribed following one of the established discourse transcriptions systems, that of Du Bois, Schuetze-Coburn, Cumming, and Paolino (1993).[6] Prior to this extract, John Laws, the host, has been discussing a hailstorm that had hit Sydney the previous evening with a Weather Bureau spokesperson. The following caller had then noted that where she was, there was still hail on the ground the next morning. Julie then follows on with this topic.

(2) Gazebo (ART COMe4 [Caller 6])

1 John; Okay?
2 Julie?
3 Julie; Morning John.
4 How are you love?
5 John; Pretty good,
6 tha<@>nk you</@>.
7 Julie; =I've still got me: ^hail in Mascot here.
8 Broke me gazebo but.
9 @@@@
10 John; Oh did it really?
11 Was it ^that heavy.

In this transcription the structure of the interaction is much more readily apparent. For example, we can see that in line 8 Julie finishes her turn with a Final Particle *but* (see Section 3.3), followed by laughter. Her prosody is final, and John responds immediately with an appreciation token *oh did it really?*.

What we are asserting in this section is that a critical aspect of creating a quality spoken component of the AusNC is using a discourse transcription system that recognizes prosodically defined

[5] <http://www.ling.mq.edu.au/shlrc/resources.htm>
[6] Unless otherwise noted, the Du Bois et al. (1993) transcription system is used throughout this paper, adapted such that within an utterance ':' indicates prosodic lengthening and '=' indicates latching. Symbols used in this paper include '@' = laughter, '(H)' = in-breath and '<F>' = forte, loud.

units as primary. To make the corpus maximally useful to a wide range of researchers, such an approach is equally essential as the quality and range of recordings, the supporting documentation and the searchability of the corpus.

2.5. Searchability

A major consideration for any corpus to be used by multiple researchers is its searchability. For the AusNC, the ability to search the documentation and transcription, as well as the recordings themselves, and in a range of ways, is a priority. For instance, allowing researchers to search by biographical details such as gender, age, and language background will allow the results to be delimited so that only speakers of interest for their particular project are shown. This is achieved relatively well in the reissue of the publicly available Mitchell and Delbridge corpus (1997-1998) where researchers can enter multiples of these values to view the sound files which fulfil the criteria selected.[7] In this reissue of the Mitchell and Delbridge corpus it is also possible to browse associated documentation relating to the project, such as the original instructions to participants and the aims of the project. The ONZE miner software[8] allows for filtering by age, gender, place of birth and the place of birth of each parent. We propose a similar situation for AusNC would be extremely useful.

Additional factors that would be useful to search by are the ability to view linguistic features removed from their context; for example, to search the entire corpus (or a large group of recordings) for all instances of a particular phoneme, such as all tokens of /b/ or /t/, or to search for certain phonemes within one sound file as well as lexical or morphological features which are more usually considered in the construction of corpora.

A number of the suggestions made here would require that the transcription and subsequent coding used in the corpus are fully searchable. The most beneficial approach would be for transcription and coding tiers to be interlinked, so that information about a single token of interest, or sequences of them, and how they fit into an overall linguistic structure, could also be extracted. This particular feature is present in *The Emu Speech Database System*, a program used primarily for phonetic analysis (see for example Bombien, Cassidy, Harrington, John, & Palethorpe, 2006).

We have provided here some examples of the ways in which a search engine would best work in a corpus of spoken language. As mentioned, there are a number of other spoken corpora which have done this successfully, and a closer reflection on these in the development stages, particularly the Mitchell and Delbridge reissue (1997-1998), and the use of open source software such as the ONZEminer tool would prove invaluable.

3. A Sample

In this section we use a sample recording to demonstrate two very different types of information that can be obtained through putting the design principles outlined in Section 2 in place.

3.1. The Example

The excerpt analysed here, which comes from a recording made by the third author in 2002, is part of an interview with ‘John’, who was 19 years of age at the time. He had lived in Melbourne all his life, attended a government school and then gone into a fulltime apprenticeship. English was the only language he spoke although he had learnt Japanese, German and French at school.

The excerpt (T=111.74 secs) was transcribed in the Du Bois et al. (1993) discourse transcription system using *Praat* (v.5.0.42). (3) below illustrates the type of transcript that can be exported from

[7] However, some factors that would be useful to search by are not included, such as age of the speakers (although in this corpus, the speakers are within a limited age range, between 16–18).

[8] ONZEminer is open source software. See http://www.ling.canterbury.ac.nz/jen/onzeminer/ for further details.

Praat.[9] Just prior to the portion of the excerpt in (3), John had described when he was hit by a car while riding his bicycle.

(3) Bike Accident (T= 56.100-66.118; lines 94-109)

1	JOHN;	.. (H) and that's my,
2		^accident[t].
3	DEBBIE;	[and] what happened?
4		Did you have to [go to hospital,
5	JOHN;	[(H) <F> Nah </F>,
6	DEBBIE;	or anything?,
7		or was it just]--
8	JOHN;	not really,
9		oh] well--
10		.. (H) I was sort of,
11		.. I didn't train,
12		for a couple of weeks,
13		.. and Mum didn't really want me,
14		to go out,
15		on the road again but,
16	DEBBIE;	Ye:ah,

In the next section we analyse the excerpt from an acoustic-phonetic perspective, whilst in the following section we analyse it from a discourse perspective.

3.2. From an Acoustic-Phonetic Perspective

A wide variety of phonetic parameters can be analysed acoustically. Some examples are segmental properties such as duration, vowel formant frequencies or voice onset time, or suprasegmental properties such as stress, loudness, or fundamental frequency (f0). In this section, we describe a suprasegmental phonetic analysis and provide results.

For this analysis, we used the ToBI transcription system for Australian English (see for example Fletcher & Harrington, 2001) to annotate pitch events, and used *Praat* (v.5.0.42) for labelling. Pitch accents (marked T*, where T=tone) and phrase-boundary tones (marked T-T%) were then measured. The combination of these categories also allowed us to label the type of tune the speaker used, for example a *high-rising tune* (e.g. H H-H%) or *continuation tune* (those ending with H-L%). As well as categorizing pitch events and tune types, we also measured the f0 of these events, to determine the overall use of pitch by this speaker.

The speaker used four types of pitch accents across the sample of speech analyzed. These, as well as their number and average f0, are shown in the table below.

[9] Transcript produced using Kendall, T. (2009) Convert from Praat TextGrid to a (Readable) Text file. Available at http://ncslaap.lib.ncsu.edu/tools/praat_to_text.php

Table 1.
Distribution and f0 of Pitch Accents

Pitch Accent	**No. (%)**	**min.**	**av.f0**	**max.**
L*	11 (12)	83.2	87.2	96.8
L+H*	10 (11)	89.6	115.6	151.7
H*	57 (63)	85.9	94.2	118.1
!H*	13 (14)	84.7	96.1	120.6

The pitch accents shown in the table above are low pitch (L*), bitonal (L+H*), high (H*) and downstepped high (!H*). By far the most common type of pitch accent used by the speaker is the high pitch accent, while the remaining three accents are used to about the same degree.

Categorization of pitch accents is relative (both to the speaker, and to each particular intonational phrase), and here we see very little difference in the average pitch of high (H*) and low (L*) pitch accents. High pitch accents are on average 94.2 Hz, while low pitch accents are on average only seven Hz lower at 87.2 Hz. However, looking at the pitch range for these two categories shows us that in fact the low pitch accents fall within a narrow range of 83.2 to 96.8 Hz, while the high pitch accents have a wider and higher range, between 85.9 and 118.1 Hz.

The bitonal pitch accent, L+H*, is a rising pitch accent, and is often used in tunes that indicate emphasis or assertion (e.g., Fletcher, 2004). As such, it is not surprising that the H* in this bitonal pitch accent measures more than 20 Hz higher on average than the simple H* accent, at 115.6 Hz.

Finally, downstepped high pitch accents, !H*, are often triggered by bitonal pitch accents. These downstepped accents are defined as high pitch accents which occur in a compressed pitch range (see for example Fletcher, 2004), and so are by nature lower than the preceding bitonal accent which triggers the compression. This is why John's average pitch for !H* accents is 96.1 Hz, much lower than the average pitch for bitonal accents. This is also slightly higher than the average pitch for simple H* accents (which are 94.2 Hz on average). Even though the pitch range is 'compressed' after bitonal pitch accents, this compression is relative, and so !H* are still somewhat higher than standard H* accents.

This analysis of pitch goes some way to explaining John's use of pitch in the sample, and also exemplifies the way that ToBI categories are used to categorize pitch. This is also shown further in Table 2, which focuses on tune endings.

Table 2.
Distribution and f0 of Phrase-Boundary Tones (Endpoint of Intonational Phrases)

Phrase & Boundary Tone	**Description**	**No. (%)**	**av. f0**
L-L%	falling	9 (16)	85.8
L-H%	(mid) rise	10 (19)	100.2
H-H%	(high) rise	6 (11)	120.8
H-L%	plateau	29 (54)	90.1

This table shows that the most common phrase-boundary tone, occurring for over half of all intonational phrases at 54%, is H-L%. This is a plateau-like tune, ending in the middle of the speaker's pitch range, and perceived as a level pitch (e.g., Fletcher 2004). This is used by the speaker primarily to indicate continuation; to hold the floor. The falling L-L% occurs at 16% of all tune endings, while the rising tunes (H-H% and L-H%) together make up 30% of all tune endings. In terms of the average f0 for these pitch events, we see results which are not surprising, and match the categories to which they belong.

Turning now to the final intonational analysis, the tune types used by the speaker are shown in Table 3. These include the last pitch accent in an intonational phrase, as well as the phrase-boundary tone combination discussed above; together making up what a listener perceives as, for example, a high rising tune (H* H-H%), a low-rising tune (L* H-H%) or a declarative tune (H* L-L%). These results overlap somewhat with those shown in the previous two tables, and are included because they give a different sense of the speaker's intonation patterns.

Table 3.
Description, no. and Proportion of Tune Types

Tune Type	Description	No. (%)
H* H-H%	high-rise	2 (4%)
L* H-H%	high-rise (from low)	4 (7%)
H* L-H%	fall-rise	9 (16%)
L+H* L-H%	rise-fall	1 (2%)
L+H* H-L%	rise-plateau	1 (2%)
H* H-L%	plateau	28 (52%)
H* L-L%	falling	8 (15%)
L* L-L%	low-fall	1 (2%)

Table 3 shows that a wide variety of tunes were used across the sample, but also that John tended to use certain tunes more than others. The simple plateau-like tune, H* H-L% was used at a level of 52% across the data, while one other plateau-like tune was also used, L+H* H-L% (making up the 54% of plateau-like tune endings discussed above). Another relatively common tune, but one used far less often, was the fall-rise tune. This is also known as a *continuation rise* (Fletcher, 2004), and occurred at a rate of 16%. The declarative tune, H* L-L%, was used at a similar rate of 15%. Other tunes were used marginally, although the high-rise (from low), L* H-H%, was somewhat more common than others.

Overall, we can say that, for tune types, the findings here for a young urban male speaker of AusE accord with the findings of Fletcher and Loakes (2006), who analysed a corpus of urban and rural speech produced by adolescent female Australian English speakers. In those two corpora, the majority of tunes used to hold the floor were mid-level or rising tunes (while declarative tunes were used more for turn yielding). While we have not carried out a full dialog act coding as in Fletcher and Loakes (2006), informal analysis indicates to us that John, who is for the most part producing monologic speech, tends to use these tunes in the same way.

Additionally, results in this section have shown how an analysis of intonational pitch events gives a different kind of profile of John's speech, using quantifiable acoustic-phonetic techniques.

3.3. From a Discourse Perspective

In approaching the excerpt from a discourse perspective, we could look to discover new aspects of interest in AusE or we could further investigate a variety of aspects of AusE that have been considered to date including the use of *yeah no* (Burridge & Florey, 2002), semi-modals such as *need to* and *want to* (Collins, 2007), and response tokens (Gardner, 2001). The aspect that we focus on here is the use of Final *but*.

The phenomenon of Final *but* in contemporary AusE has been documented in Mulder and Thompson (2008) and Mulder, Thompson, and Penry Williams (2009). In these studies evidence is presented that there are two types of Final *but*, which are termed Final Hanging *but* and Final Particle *but*. Both types of Final *but* are shown to have two essential features: 1) they end an IU and 2) they end a turn. It is proposed that what differentiates the two types is that the Final Particle type has progressed through a grammaticization continuum to become a 'fully developed' final discourse particle, parallel to that of *though* (Barth-Weingarten & Couper-Kuhlen, 2002). Furthermore, it is observed that while both American English and AusE have Final Hanging *but*, only AusE has Final Particle *but*.

Focusing on Final *but* in AusE, Mulder et al. (2009) make the case that *but* as a final particle provides further evidence of the mixed origins of AusE (Bradley, 2003; Leitner, 2004) and that while it is often claimed that AusE has few distinctive grammatical features, in particular as distinct from British and American Englishes (cf. Lass, 1987; Newbrook, 2001), it can be seen as a distinctive feature of AusE. In support of this they establish through survey data that its usage in AusE differs from that in American English. Lastly, they argue that, as evidenced in written dialogue, emails and text messages, Final Particle *but* has social meaning and can index "Australianness".

Looking more closely at the difference between the two types of Final *but*s, these two studies show how prosody, turn organization, and speaker interaction in contemporary conversational data indicate that with a Final Hanging *but* there is a clear implication left 'hanging', such that the clause ending with *but* is open to being interpreted as a concession, with the claim for which it is a concession only implied. That is, this *but* tells the hearer that there is an implication, and invites the listener to infer what it is and to continue the interaction appropriately given that implication. In contrast, instead of leaving an implication 'hanging', with a Final Particle *but* the semantically contrastive material is supplied in the IU ending with the final particle.

To illustrate this difference, consider example (3) in Section 3.1 again. John's utterance in line 15 ends in a Final Hanging *but*, where the implication left hanging is a contrasting idea for the hearers to infer, perhaps something like 'but I did anyway'. Debbie's *yeah* in line 16 indicates that not only is she orienting to this point as a turn-transition point, but that she has gotten the implication and is affiliating herself with it. On the other hand, in example (2) in Section 2.4, Julie finishes her turn in line 8 with a Final Particle *but*, conceding that even though the hail broke her gazebo, in a drought, precipitation in any form is a positive. There is no unstated, 'hanging' implication following the Final *but.* Rather, the semantically contrastive material is supplied in the IU ending with the Final Particle *but.* That is, the Final Particle *but* closes a construction which conveys semantically contrasting content. The difference in these two types of final *but*s is further supported by prosodic evidence: In (3) the prosody of the IU with Final Hanging *but* accords with its function of leaving an implication 'hanging' as indicated by the comma following *but*, whereas in (2) the prosody of the IU with Final Particle *but* correlates with an interactional 'closing', as indicated by the full stop.

An important element to draw from this discussion is the fact that the discourse transcription system utilized in examples (2) and (3) allows us to readily locate and then 'see' and 'hear' the interaction, giving insight, in this case, into the use of Final *but* in AusE.

4. Suggestions for Putting Together a Spoken Component

Learning from the advantages and disadvantages of the International Corpus of English – Australia (ICE-AUS)[10] and the Monash University Dimensions of Australian English corpus (DAusE) and also from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al., 2000, 2003; Du Bois & Englebretson, 2004, 2005) and the spoken section of the British National Corpus (BNC) (Burnard, 2000), the following is suggested as a possible plan for putting together a spoken component of the AusNC which would include all the design considerations discussed in Section 2.

Following the BNC and SBCSAE models, the first step would be to actively seek people from different regional and social backgrounds who are interested in being involved in data collection. The responsibility of recruiting these people could be divided up between interested ALS members across Australia with quotas to prevent over representation in terms of region, socioeconomic status, age, gender, etc.

As was done with the DAusE corpus, these 'recruits' would then be interviewed not only to establish all the background information needed, but also to allow various tasks to be completed (e.g. reading passages, wordlists, etc.). The interviews could be conducted either by the ALS member or a postgraduate student in line with an explicit interview schedule which all interviewers would use.[11]

[10] <http://www.ucl.ac.uk/english-usage/ice/iceaus.htm>

[11] There is no reason why the interviewee and interviewer need to be strangers. As long as their relationship is noted it could be a student, family member or friend of the interviewer.

The DAusE, SBCSAE and BNC project all supplied audio equipment. It is suggested that at the end of these interviews the interviewees be given recording equipment to take away with them after brief training. This equipment would be set to record according to the requirements noted in Section 2.1. The interviewers would be contactable by phone if the interviewees should need any further technical support. Of course, the recruits would need to be highly motivated and keenly interested in the project to ensure willingness to take on such responsibilities.

As done in the BNC project, the interviewees would then record a handful of various interactions over the following week. These could be specified, such as a family meal, a meeting, an interaction with a friend, or left open. However, it should be stressed to the interviewees that the interactions should not be especially planned for the purpose of recording. The original interviewees would supply all people present at the time of the recording with consent forms and plain English statements and be clear about issues of ethical research. They would be responsible for entering information on each of the interactions they record and for asking and ensuring the other participants fill in an online form for themselves, after supplying them with the web address and the code/name for the interaction in which they participated. The interviewees would then return the equipment, consent forms and recordings in person a week later and at that time would be given the chance to make any comments or ask any questions they wanted to. Time would be allotted for the interviewers to speak to the interviewees and to ask any questions about the (reviewed) information supplied via the online forms.

All recordings would be submitted to a review panel to ensure only high quality recordings with complete information are included in the corpus. For each original interviewee, the panel would assess the quality of the data collected and choose the best recordings made. They would then group-rank the cases, and according to the resources available, decide on a cut off point. Ideally, for each case above the cut off point, the initial sociolinguistic interview and at least two other interactions which include this speaker would be selected for transcription and inclusion in the corpus. This allows for the number of initial sets of recordings included to reflect the budget available for transcription, other recordings being kept to be added as further funding is received in the future.

The selected recordings would then be transcribed using an established discourse transcription system such as that of Du Bois et al. (1993). As well each recording would be transcribed at the phonemic level.[12] As part of the transcription process there would need to be a training module which sets out the basic categories, symbols, and conventions used in the transcription system. For each symbol, a brief explanation of usage would need to be given along with an illustrative example, both audio and transcribed, drawn from actual interactions.[13] Each audio recording would be transcribed independently by two people trained in the agreed discourse transcription system and inter-transcriber agreement checked.

If funding is inadequate for complete discourse and phonemic transcriptions of all sets of recordings above the cut off point, individuals who use the data might be able to add transcribed sections or levels of detail via multiple author technology. For example, a phonetician looking at realisations of /t/ marks all the /t/ segments, a discourse analyst adds a transcription of several different one-minute segments, and in this way the database grows. Files deemed of appropriate quality by the review panel but not yet included could be available to researchers upon request if they provided any transcription they completed. These could then be included for inter-transcriber agreement checks, and form part of the transcriptions added to the corpus.

To summarize, the process we have outlined here consists of the following stages:

1. Speaker recruitment and selection (balanced with respect to age, region and social factors)
2. Initial speaker interview
    - Sociolinguistic interview in line with an interview schedule, yielding
        - Extract for corpus - interview setting

[12] Our understanding is that there is software available that can provide automatic phonemic transcription (based on standard pronunciation). The resulting transcription would then only need to be altered to reflect actual pronunciation (i.e. on a phonetic tier).

[13] An initial set of transcribed recordings could be prepared as a model, with subsequent additions to the corpus serving as further models.

    - Speaker documentation
  - 'Other tasks' such as reading lists
3. Speaker Training
  - Obtaining informed consent
  - Using recording equipment
  - Collecting supporting documentation
  - Discussion of interactions to record
4. Speaker record a handful of various interactions over the following week
5. Speaker follow-up
6. Audio extract submission, selection and ranking by a review panel
7. Transcription process
  - Training module (covering basic categories, symbols, conventions, examples)
  - Independently transcribed by 2 transcribers and inter-transcriber agreement checked
  - Acceptance by a review panel
8. Set of recordings, transcriptions and supporting documentation added to the AusNC

5. Conclusion

In this paper we have argued for the importance of creating a quality spoken component of the AusNC. We have suggested that to support a wide range of existing and new research areas on AusE, the data must be of high quality, incorporating best practice processes. Toward this we have given a set of design principles with detailed specifications in terms of recording requirements, the range of recordings to include, the scope of supporting documentation necessary, the level of transcription needed and the types of searchability that are advantageous. In support of these design principles we have used a sample recording to demonstrate the value of the information which can be obtained through putting such design principles in place. Finally, we have presented what we believe is a viable proposal for generating a quality spoken component of the AusNC.

References

Barth-Weingarten, Dagmar, & Elizabeth Couper-Kuhlen. (2002). On the development of final though: A case of grammaticalization? In Ilse Wischer & Gabriele Diewald (Eds.), *New Reflections on Grammaticalization* (pp. 345–361). Amsterdam: John Benjamins.

Bombien, Lasse, Steve Cassidy, Johnathan Harrington, Tina John, & Sallyanne Palethorpe. (2006). Recent developments in the Emu Speech Database System. In Paul Warren and Catherine I Watson (Eds.), *Proceedings of the 11th International Conference on Speech Science and Technology* (pp. 313-316). Auckland: Australasian Speech Science and Technology Association, Inc.

Bradley, David. (2003). Mixed sources of Australian English. *Australian Journal of Linguistics, 23*(2), 143-150.

Burnard, Lou (Ed.). (2000). *Reference guide for the British National Corpus*, available at http://www.natcorp.ox.ac.uk/docs/userManual/

Burridge, Kate, & Margaret Florey. (2002). 'Yeah-no he's a good kid': A discourse analysis of *yeah-no* in Australian English. *Australian Journal of Linguistics, 22*(2), 149-171.

Chafe, W. 1994. *Discourse, Consciousness, and Time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Collins, Peter. (2007). Modal wars: Some ascendant semi-modals in Australian English. In Jeff Siegel, John Lynch, & Diana Eades (Eds.), *Language description, history and development: Linguistic indulgence in memory of Terry Crowley* (pp. 197-208). Amsterdam: John Benjamins.

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

Du Bois, John W., Wallace L. Chafe, Charles Meyer, and Sandra A. Thompson. (2000). Santa Barbara Corpus of Spoken American English, Part One. Philadelphia: Linguistic Data Consortium.

Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, and Nii Martey. (2003). Santa Barbara Corpus of Spoken American English, Part Two. Philadelphia: Linguistic Data Consortium.

Du Bois, John W. and Robert Englebretson. (2004). Santa Barbara Corpus of Spoken American English, Part Three. Philadelphia: Linguistic Data Consortium.

Du Bois, John W. and Robert Englebretson. (2005). Santa Barbara Corpus of Spoken American English, Part Four. Philadelphia: Linguistic Data Consortium.

Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming, & Danae Paolino. (1993). Outline of discourse transcription. In Jane A. Edwards & Martin D Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 4–89). Hillsdale, NJ: Lawrence Erlbaum.

Edwards, Jane. (2001). The transcription of discourse. In Deborah Schiffrin, Deborah Tannen, & Heidi E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 321-348). Oxford: Blackwell.

Edwards, Jane A., & Martin D. Lampert. (1993). *Talking Data: Transcription and coding in discourse research* Hillsdale, NJ: Lawrence Erlbaum.

Fletcher, Janet, & Johnathan Harrington. (2001). High rising terminals and fall- rises in Australian English. *Phonetica, 58*(4), 215-229.

Fletcher, Janet. (2004). *Prosody.* Lecture notes from ALTA Summer School, Macquarie University, Dec 8-10. available at www.alta.asn.au/events/altss2004/course_notes/ALTSS-S-Fletcher-Prosody.pdf

Fletcher, Janet, & Deborah Loakes. (2006). Patterns of rising and falling in Australian English. In Paul Warren & Catherine I. Watson (Eds.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 42-47). University of Auckland, New Zealand.

Gardner, Rod. (2001). *When listeners talk: Response tokens and listener stance.* Amsterdam: John Benjamins.

Holmes, Janet, Bernadette Vine, & Gary Johnson. (1998). *Guide to the Wellington Corpus of Spoken New Zealand English,* available at http://khnt.hit.uib.no/icame/manuals/wsc/INDEX.HTM

Ladefoged, Peter. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques.* Malden, MA: Blackwell Publishers.

Lass, Roger. (1987). *The shape of English.* London: Dent and Sons.

Leitner, Gerhard. (2004). Australia's many voices: Australian English — the national language. Berlin: Mouton de Gruyter.

Mitchell & Delbridge corpus (1997-1998). The Mitchell and Delbridge Tapes: 1998 Reissue through the School of Electrical and Information Engineering, Faculty of Engineering, and the Language Centre of the Faculty of Arts, The University of Sydney, available at http://www.ee.usyd.edu.au/speech/MDdb.html

Mulder, Jean, & Sandra A. Thompson. (2008). The grammaticalization of *but* as a final particle in conversation. In Ritva Laury (Ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions* (pp. 179-204). Amsterdam: John Benjamins.

Mulder, Jean, Sandra A. Thompson, & Cara Penry Williams. (2009). Final *but* in Australian English conversation. In Pam Peters, Peter Collins, & Adam Smith (Eds.), *Comparative grammatical studies in Australian and New Zealand English* (pp. 337-358). Amsterdam: John Benjamins.

Newbrook, Mark. (2001). Syntactic features and norms in Australian English. In David Blair & Peter Collins (Eds.), *English in Australia* (pp. 113-132). Amsterdam: John Benjamins.

Ochs, Elinor. (1979). Transcription as theory. In Elinor Ochs & Bambi Schieffelin (Eds.), *Developmental Pragmatics* (pp. 43-72). New York: Academic Press.

University College London, Dept. of Speech, Hearing & Phonetic Sciences (2009) *Equipment for audio recording of speech,* available at http://www.phon.ucl.ac.uk/resource/audio/recording.html

Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages

edited by Michael Haugh, Kate Burridge, Jean Mulder, and Pam Peters

Cascadilla Proceedings Project Somerville, MA 2009

Copyright information

ISBN 978-1-57473-435-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Mulder, Jean, Cara Penry Williams, and Deborah Loakes. 2009. Developing a Quality Spoken Component of the Australian National Corpus. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 61-73. Somerville, MA: Cascadilla Proceedings Project.

or:

Mulder, Jean, Cara Penry Williams, and Deborah Loakes. 2009. Developing a Quality Spoken Component of the Australian National Corpus. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 61-73. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2289.