elpublishing

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 5*. Editor: Peter K. Austin

Data - but data from what?

RUTH FINNEGAN

Cite this article: Ruth Finnegan (2008). Data - but data from what?. In Peter K. Austin (ed.) *Language Documentation and Description, vol 5*. London: SOAS. pp. 13-28

Link to this article: http://www.elpublishing.org/PID/061

This electronic version first published: July 2014

Data – but data from what?

Ruth Finnegan

1. Preamble

I was once confident of what 'language' was, where its boundaries lay, and hence what might count as data for documenting it. But I am no longer at all sure. Nor am I clear where information about a given language should best be found, or how and by whom a language should be documented. My uncertainties are founded in my own puzzles over the many years that I've worked, mainly as an anthropologist, on aspects of unwritten literature, performance and communication, based both in comparative reading and fieldwork in Africa and Britain.[1] Within that limited experience I find that issues I have found myself confronting are surprisingly relevant for the questions posed in this issue of *Language Documentation and Description*.

What I offer here are merely some informal reflections, not any pretence of a scholarly or theoretical disquisition.[2] I write not as a linguist nor as someone with any expertise in endangered languages, but merely about my experience of becoming increasingly doubtful of my initially confident assumptions about just where in the great spectrum of human communicating and expression we are to find 'language'.

2. I once knew what language was …

My first degree was in Classics – Greek and Latin. At that point I was pretty clear what 'language' was – or rather, I didn't need to know because it seemed self-evident. It was what came in written texts.

Written texts were the prime sources that had come down to us from classical antiquity, transmitted (mainly) through the manuscript tradition and

[1] Documented in Finnegan 1967, 1970, 1977, 1988, 1989, 1992, 1998, 2002, 2007.

[2] The editor has encouraged a personal tone for this article (hence, with apologies, the many references to my own work), so I have not clothed it with the conventional apparatus of systematic citations throughout. But since my personal experience is of course interrelated with changing and contending approaches to language and communication let me mention that works I have at various times found especially illuminating include Austin 1962, Bakhtin 1986, Bauman 1977, Bauman and Sherzer 1989, Bauman and Briggs 1990, 2003, Dalby 1999/2000, Duranti 2004, Hanks 1996, Harris 1987, Harris and Wolf 1998, Hodge and Kress 1993, Hymes 1977, Robinson 2006, Tracey 1999; also most recently several chapters (noted in the references) in Gippert et al. 2006. Some issues touched on here are considered in more fully-referenced framework in Finnegan 2002, 2007.

Ruth Finnegan 2008. Data – but data from what? In Peter K. Austin (ed.) Language Documentation and Description, Vol 5, 13-28 London: SOAS.

with, of course, no audio records of speech. The texts we read and studied were wonderful and enriching, covering a wide range of genres – literary, historical, epistolatory, oratorical, lyrical and much else. Both drawn from and supporting this corpus of texts was the extensive apparatus of vocabulary, of grammar and of syntax, all once again presented in writing in the form of dictionaries of words (usually offering equivalencies in some European language) and accounts of grammatical and syntactical rules. The written words, organised in the correct classic formulations – that was ultimately what language consisted of.

This emphasis on the textual and written was not totally unqualified. Archaeology – the study of material remains – played a part, and some scholars went beyond the printed page to read aloud a Catullus love poem or engage with the acoustic dimensions of Greek lyric meters. There was an established tradition (though not within the examination curriculum) of live performances of Greek plays or of reading Homer aloud. But the paradigm was indubitably of the centrality of written text both as the object of what was studied and the medium in which such study was appropriately expressed.

From this viewpoint, documenting a little known language would entail finding and pinning down its essential constituent: texts that could be read, analysed and form the basis for identifying underlying rules. The texts might have to be snared by transcribing spoken words into writing. But ultimately those resultant scripts, together with a similar scholarly apparatus as for classical languages, would form the necessary data of documentation. Language was capturable and realised in the communication technology dominant in the mid-twentieth century and earlier – writing. – and it was ultimately there that the data could be found.

3. … but fieldwork shook this up

Things began to look different when, as a postgraduate, I embarked on anthropological studies, followed ineluctably by my first piece of fieldwork. This was in the early 1960s among a people called the Limba, in northern Sierra Leone. My focus came to be on their stories and story-telling, an interest which followed on well from my enthusiasm for literary texts in my earlier studies. I was hugely impressed by the many story-telling performances I experienced there and wanted to make that aspect of Limba culture the central core for my thesis and subsequent work.

My initial presupposition was that the way to study these stories – and most certainly the way to present them in my doctoral dissertation – was to capture them as written text. That after all, I assumed, was where their true reality lay and the medium in which I and other scholars possessed the

necessary analytic tools. There seemed no other way, really, to properly pin them down for scholarly study. Some of the stories I transformed quickly into script by taking them down from dictation. Many others I recorded on one of the (relatively) portable tape recorders then available. The obvious next step was to transcribe from tape into written lines on a page in similar format to the classical texts I and others were accustomed to. My thesis could then take the familiar form of introductory background and analysis followed by the key data – parallel texts in Limba and English translation. It consequently ran to three large volumes (I still remember their weight as I lugged the required three copies of each through Oxford by bicycle, then up the steps to the examinations schools). I assumed – as, apparently, did my examiners – that the substantive data, the corpus of texts, had to be there in my presentation.

But there was a problem. I had been greatly struck by the richness and subtlety of these narrations, and in my thesis tried to convey something of their artistry. .And yet – that had somehow melted away in the stories I presented. At one point, trying to demonstrate why I was so enthusiastic, I showed one of the texts to a friend from my classical days expecting him to be impressed. He read through and rejoined – politely – 'Oh yes, another of those charming African animal tales', to my mind missing all its wonders.

The point is of course only too obvious, though it had taken me some time to fully appreciate it. The reality lay in the performance. It was this that the written texts had failed to capture. They missed the subtle characterisations, the drama, the way the tellers used volume, pitch, tempo, repetition, emphasis, dynamics, silence, timbre, onomatopoeia, and a whole plethora of non-verbal indications to convey humour, pathos, irony, atmosphere … . The written forms did not replicate the ideophones that peppered the tellings – those vivid little mini-images in sound and more than sound. The unilinear textual layout could not give the many-voiced interaction and co-construction by the audience as they joined in songs led by the narrator and reacted with horror or laughter to key turns in the tale. Nor did it convey the common Limba practice of picking out one among the audience as the 'replier' – someone to give special support, prompting, echoing and, where needed, exaggerated reactions and response. Compressing this multidimensional and multi-participant performance within the narrow one-voiced medium of writing was to miss its substance.

I soon discovered that similar patterns were found elsewhere – obvious once you look, but for long concealed from me (and others) by the presupposed centrality of written text. The study of oral poetry, performance, and 'oral literature' more generally hammered home the same point. Both in Africa and further afield those creating performed literary art deploy not just writable words but a vast range of non-verbalised auditory devices of which those conventionally captured in written text, such as rhyme, alliteration and

rhythm, are only a small sample. The wondrously varied expressive resources of the human voice are exploited for multifarious delivery modes, varying with genre, situation or performer: spoken, sung, recited, intoned, shouted, whispered, carried by single or multiple or alternating voices. Not just in far away places but in the spoken and sung forms nearer home too, there turned out to be near-infinite combinations of vocal expression and auditory resources of which most escape from view on the written page.

I had to conclude, then, that the core lay not in written text after all but in the performance. Ant that included the setting, the delivery, and not just the 'lead' speaker but the full range of participants. All this showed up the contentious nature of my earlier 'language-as-written-text' model. This was reinforced by on-going trends in the study of verbal expression, among them the performance-oriented approaches and ethnography of speaking in folklore and anthropology – stressing performance and process rather than text and product – as well as more recent developments in linguistic anthropology, sociolinguists, and performance studies. At the same time interdisciplinary interests in oral performance and in 'orality' more generally have been flourishing, opening up a new vision of the nature of human communication and expression previously concealed by the focus on the written.

This then turned me towards seeing language as ultimately something spoken, performed, oral. It no longer seemed to be existent essentially in written text but in active performance and interaction. And if so, language documentation would have to be approached very differently than from the familiar written-text perspective. For it would have to focus on audio, not just written, materials, and to include records and analyses of oral performances and (where relevant) their multiplicity of overlapping participants. Such data would not only count, but be essential.

4. Doubts and complexities

Acknowledging the limitations of a written-text model of language is perhaps by now scarcely problematic. Audio recordings are nowadays widely accepted as a regular (though perhaps not universal) part of serious language documentation. I would like to add two further comments however about the implications.

First, I begin with a qualification. The move away from the written to the 'oral' sometimes jumps to the opposite extreme, envisaging the spoken as somehow the bedrock, natural, traditional, to be set against the artificial imposition of writing. A seminal western myth sometimes lurks behind this, constantly challenged but also constantly recycled. This posits a fundamental opposition between two mutually exclusive types of social and cognitive

organisation: the one literate, rational, scientific, civilised, western, modern, the other communal, emotional, non-scientific, traditional, primitive – and oral. This has underpinned the trend to mystify 'orality' and the 'oral' as if something distinctive and separate: characteristic of a culture belonging prototypically to the 'them' of far away or long ago and one in which writing, even if in certain respects present, is intrinsically alien (and to be ignored). This is a set of assumptions I have long found myself struggling against and one which no doubt also crops up – controversially – in certain approaches to language documentation.

In other ways however I think that recognising the oral and performed dimensions of language has not been taken far enough. The vocabulary to capture the amazing use of voice with its huge range of subtleties is relatively little developed, and the sonic elements of language still often sidelined. But if we are to document the auditory practice of language then the data to count would need to cover not just rules about phonetics, word forms or (limited elements of) prosody but its active sonic realisation in such features as, for example, pacing and speed, volume, pitch, melody, rhythm, onomatopoeia, voice quality, timbre, mood, mix with other voices and sounds – or silences – distancing, vocalised sounds like sobs, sighs, or laughter – and so much else.. Data about tone or prosody would have to include not just smaller units like words, phrases or sentences but also the sonic patternings of larger chunks and of speech genres more widely. It's true that such elements sometimes get mentioned under the head of 'paralinguistic' or 'extra' linguistic elements – but in an oral-performance model of language these are not supplementary extras but intrinsic. A Martian anthropologist might well be puzzled by a demarcation which included some auditory elements in the delineation of language but excluded others which can equally form part of both the conventions and the unique personality communicated through human vocal utterance.

So though the importance of audio features may now be increasingly taken for granted in documenting languages, helped by the audio technologies which now facilitate the recording, storage and accessing of such data, has this yet been fully followed through? Documenting the oral is inevitably enormously complex, nor, despite the wizardries of modern technology, have we really developed adequate techniques, vocabularies or perhaps concepts to fully capture and analyse these inevitably more fleeting and temporal performed features? Small wonder perhaps that the written model of language is so extraordinarily persistent, with its implicit suggestion that data doesn't quite 'exist' until it is reduced to, transcribed as, transformed into, or analysed through the spatial solidity of writing and print. As Hodge and Kress well put it 'The distinctive resources of spoken communication which are not transcribed are eliminated from linguistic theory' (Hodge and Kress 1993: 11). Even when we accept a view of language as sounded and performed do

we still fall comfortably back into a model in which the true reality – and the key data – reside in visually written textualisations rather than vocal enunciation?

5. Cognitive models?

My Limba fieldwork brought me face to face not just with story-telling performance but also with the active way that Limba speakers used vocal utterances to *do* things. This, I gradually discovered, ran counter to a further implicit model of language that, if only in a vague and muddled way, I'd also had at the back of my mind.

This was a set of somewhat contradictory and elusive assumptions, which could indeed be split apart but which nevertheless tended to come together in a kind of general mindset which I'd sum up under the label of 'cognitive'. Basically I pictured language as something essentially mental, rational, decontextualised. Language was to do with mind and meaning, and its central function was referential. Artistry and rhetoric were secondary embellishments in contrast to its core prose and information-bearing elements. Language might or might not constitute an independent rule-governed system existing autonomously in its own right – I vaguely assumed that it did – but it certainly had a structure that could be abstracted from the messiness of context, usage and social action or experience.

Of course I should already have known that that wasn't the whole story, both from my own experience and from my encounter with the multiplicity of classical genres. Even so I was still somehow steeped in that set of preconceptions. It had been reinforced in part by the legacy of logical positivism still influential in my undergraduate years at Oxford (though tempered by Austin's lectures on 'performative utterances' which were much to influence me subsequently). More radically, as I came to realise, it was a continuance of an ideology powerful in western thought over several centuries which asserted the rationality of language and its relation to science, objectivity, civilisation, literacy and, ultimately, the achievements of the west.

In some ways it was a serviceable model for a field situation. My language learning had indeed initially relied on the presupposition of a systematic vocabulary and grammar that I could learn independently of the pressures of spoken situations. There was a short missionary-compiled Limba dictionary, a couple of translated gospels, and two short articles based on elicited data by a SOAS linguist (Jack Berry), all of which I found hugely helpful. They fitted both my preconceptions about the systematised and meaning-carrying nature of language and where to find data about it, and my conviction that meaning could be conveyed cross-culturally and out of context. Language as the

repository of meaning offered the potential for its 'translation', a channel by which minds could be brought into contact across space and time. It was through language that the Limba stories could be transported to others, something which I indeed aspired to do through my verbal translations.

My aim was not to document language as such, whether that of Limba speakers or any others. But if it had been I would doubtless have started from the assumption that the core data would be found in the information-carrying forms, in 'plain prose' sentences and the logical structure underlying them; also that I would have to produce clear translations and word-equivalences to enable the direct transference of meaning from this lesser-known culture into some accessible European language.

But greater experience of Limba life somewhat undermined that set of preconceptions. I could not really miss the way Limba speakers used speaking as organised action and performance rather than, or in addition to, for conveying meaning. They used language to *do* things rather than just *describe* them: to recognise and forge relationships, ratify contracts, issue orders, assert a position, strike an attitude, show off as performer.

Further, in some interchanges, and even in some Limba stories, the cognitive 'content' as it were - the meaning I had assumed I could transfer – was *not* after all the only, or in some cases even apparently the most important, element. I think for example of one ridiculous short story I recorded about a fictional character called Daba, an incorrigible snuff-taker. All that happened was that Daba went round the local chiefs badgering them to give him vast quantities of snuff, then finally over-reached himself by taking a huge sniff and falling down dead: nothing to it really. And yet this was hugely successful with the audience, who were rolling with merriment. It was told by one of the most admired local tellers, and among the liveliest narrations I encountered, subtle as well as hilarious. Its success lay not in its plot but in the teller's brilliant performance and the audience's active co-creation and singing as Daba sniffed and sniffed again, also in the narrator's skill in exploiting their shared knowledge of local personalities, satirised as Daba goes the round of the chiefs, and of the ludicrous way some people carry on, held up to mockery in Daba's absurdly extreme personality.

I had also rather assumed that in focusing on stories I had managed to select a core linguistic genre: narrative, close to 'ordinary speech' and thus somehow basic in a way that their songs and more overtly 'artistic' behaviour were not. I tacitly congratulated myself on that, feeling it took me direct into something primary about their language. But I came to acknowledge that story-telling was no more nor less 'natural' than any other genre. It too had its own speech conventions. Nor was there anything special about either narrative or (so-called) 'prose' that gave them any more seminal or objective status than

anything else. All cultures, I had to accept, recognise a variety of 'speech genres', as Bakhtin (1986) famously has it, each with their own poetics.

Not that everything about a cognitive view of language seemed wrong to me. But both from fieldwork experience and more comparative work on literacy and communication media more generally I became doubtful how far that set of preconceptions could adequately illuminate either the Limba experience or human culture as it was realised in practice. And if so, the data necessary for documenting a language would seem to involve not primarily matters to do with 'its' abstract linguistic system, translateable cognitive meaning or supposedly 'primary' forms such as narrative or conversation, but data from and about the full range of recognised genres. It would have to cover the near-unending and diverse ways people used and enacted language, for art, action, reflection, play or whatever. An impossible project? But might aiming at anything less risk invoking a seriously incomplete model of language?

6. Where are the boundaries?

And amidst all those puzzles, I have also become unclear how to divide language from other (but are they other?) modes of human expression.

One uncertainty that has particularly dogged me is the relation between music and language. Some cases are perhaps clearly one or the other, but where if anywhere does the line come?

Take intonation. I think I had originally assumed that this was to do with individual words or sentences and as such a relatively accepted, if limited, dimension in some (perhaps not all) approaches to language. Thus in the Limba stories I recorded I took it that intonation was effective in particular phrases and how they were delivered but not of much interest in the narration more widely. But I changed my mind when, unexpectedly, I was played an audio recording of a Mossi story from some hundreds of miles away, in a very different West African language. I knew no Mossi so listened to the sounds. I was amazed to hear familiar intonational and rhythmic patterning in long passages of the telling. It could have been a Limba performance. I had not noticed before how part of the characterisation of the genre was its sonic shaping.

A similar point emerged from the comparative study of oral poetry. Not only were there many varieties of rhythmically and sonically patterned delivery, delineating both particular generic conventions and unique performance attributes, but some poems were performed in a way that meant they could equally well be described either as 'sung poetry' or as 'vocal music' – or, indeed, as 'song'. In these performed genres, enacted by single or

multiple voices, sometimes instrumentally embellished too, should I really be endeavouring to separate 'linguistic' from 'musical' elements, and if so how? The same applied in the urban music making I studied in both Fiji and England - tearing apart the 'song texts' (as, like many other scholars, I often in fact found myself doing ...) was in practice to mangle the songs' reality as they were actually delivered.

It is true that in some cultural contexts a music/language division seems self-evident. In the European high art song-tradition of 'text-setting', words and music are indeed in a sense separated. But it has in fact been urged for some time that the apparent distinction between language and music would be better represented as a continuum rather than dichotomy [3] In practice it is near impossible to drive a clear wedge between the multifarious modes of vocal expression - speaking, intoning, chanting, recitative, melodic singing, and so much else. Ethnocentric too, given that the classifications of different cultures vary. Even in western experience the classical Greek *mousiké* originally had a different coverage from the modern 'music', for it encompassed what we would now differentiate as music, poetry and dance, while the mediaeval *musica* covered spoken as well as sung performance, with little idea, apparently, of words and music as 'separate expressive media that one could choose to unify or not' (Treitler 2003: 47). Indeed even in modern times can one really divide up the music and the language of vocal performance, whether T. S. Eliot declaiming his poetry, Edith Sitwell chanting her 'Façade', a fine reading of a Shakespeare sonnet or a contemporary rap or dub performance? All these resonate through the sounding voice as people deploy an unending wealth of sonic resources in their vocal utterances.

So should the melodic and rhythmic qualities of performed vocal utterances - what some might separate out as 'music' - be relevant data for language documentation? How far to include them must depend on where and whether we are prepared to draw a boundary between music and language – and that¸ it seems, is far from unambiguous or culturally neutral.

Problems about boundaries do not just relate to audition, as is sometimes assumed from too enthusiastically embracing the concept of 'oral' / 'orality'. As I learnt from watching Limba narrators, performers can also draw strikingly on visual resources. Not just in Limba contexts – the setting which first most directly alerted me – but, I now realise, in communication more generally, people make use of gesture, facial expression, eye glances, bodily orientation, demeanour, movements, material artefacts. To learn a language fluently includes mastering the appropriate visible actions belonging to particular genres or situations. So where do you draw the line?

[3] See List's influential article (1963), also more recently Feld and Fox 1994, Banti and Giannatasio 2004, Finnegan 2006.

The question is raised particularly by gesture. In many standard approaches this is set apart from language. But recent studies of the intimate ways gestures are systematically co-ordinated with speech [4] have raised the question of whether the boundaries of language should be widened to include them. Here again modern communication technologies have expanded our capacity to capture - and thus notice - the significance of moving images, endowing them with a solidity concealed when we limit ourselves to script-based tools. So is it justifiable or not to claim that any language could be fully documented without data on the uses of gestures?

Once we go beyond models of language as centred on written text or on abstract or cognitive systems, and consider practice and performance, it also becomes inescapable that human communicating is commonly multisensory. As well as audio and visual elements – many-sided as these already are - tactile and somatic elements may be in play too, as in the danced and embodied movements that characterise some genres and performances. The physical setting and spatial arrangements can carry their resonances too, not least in the multisensory proxemic relations between participants. Multimodality may be more to the fore in some genres than in others. But where it *is* a feature, should we screen out such data by implicitly invoking a model of language where such dimensions do not really count?

I have also been intrigued by the diverse ways in which representation in other modes or materials – not just music or gesture - can work alongside or be variously linked or paralleled with speech. Pictures, sculptures, drumming, sign languages, tactile tools, web representations – there are a host of complex interrelationships. These too may in any given case be closely tied into verbal usage and arguably count among the data that should count. But they are likely to work – and be conceptualised - differently in different situations and cultures, and a link seen as self-evident in one setting to be highly problematic in others.

In the case of the visible marks labelled as 'writing' there might seem no argument – these surely are inextricably tied into 'language'. But the comparative study of literacy has raised two issues for me. First, insofar as we do recognise close speech-writing ties, then data about this particular form of material representation is indeed relevant in documenting the language. At one point the established presupposition - by which I was implicitly swayed during earlier fieldwork but revised when I came to study literacy more directly – seemed to be that whereas written forms were pretty basic in 'developed' languages, elsewhere writing was somewhat intrusive and alien. If so perhaps it did not really count among the authentic data for some kinds of languages? On the other hand if written forms are in fact current now,

[4] See for example Haviland 2004, Kendon 2000, McNeill 2000.

should they not be considered relevant for the present linguistic situation? I (and many others) would probably now say that they should, and would also want to include in the data not just oral/written contrasts but their interactions and, perhaps, interpenetrations. But then that, of course, is again to make particular assumptions about the scope of language.

That leads to a second question. In western contexts it has seemed self-evident that language can reasonably be identified, in broad terms, as speech and writing (and in particular alphabetic writing): the link seems a natural one. But to take for granted that these two media have a given one-to-one equivalence is perhaps cross-culturally problematic. And if writing is to count, then what about the other modes and media that are in one way or another closely linked into speech – pictorial, material, tactile or whatever, varying in different cultures? Should they too be considered as potentially relevant data?

So does language turn out to spill across *all* the resources that human beings so wonderfully exploit in their communication and expression alongside, or intertwined with, speech? For practical purposes the boundaries have to be drawn somewhere no doubt. But to do so is unavoidably to take up a particular stance and thus become liable to criticism as incomplete, lop-sided or ethnocentric. For *wherever* they are drawn is to make debateable assumptions about the nature and limits of 'language'.

7. How and by whom should data about 'a' language be documented?

My puzzles about language also extend into queries about who or what should be involved in providing and collecting the data.

I can bypass the well-worked issue of just where the boundaries of 'a' language can be set, since the older picture of unitary and exclusive languages seems to have been replaced by a more realistic awareness of relativity and diversity. But I would like to comment on the commonly-used and partly analogous term 'speech community'. In many ways I find this concept helpful, especially for its focus not on abstract systems but on people and usage. But 'community' is itself a controversial and elusive concept. It raises questions of who demarcates and draws its edges and whether these are defined in terms of, say, location, identity, perception (and whose perceptions?). It has long been tempting to see something dubbed as a 'community' as homogeneous and bounded – when in practice it might equally well be heterogeneous, made up of perhaps warring interests, without clear boundaries and by no means necessarily permanent. There is perhaps an additional pull to romanticise 'communities' that consist of people who can be thought of as somehow other – minorities, far away, long ago, or,

alternatively, in some way an issue on the political horizon. Sometimes the term evokes that still emotive image of the homogenous, unchanging, and romantic past. In my Limba fieldwork I was less critical than I should have been of the temptation of positing generalised 'traditional' patterns - even though I knew there were differences in different areas and that 'the Limba' had been demarcated by colonial administrators and others as speaking one language (and 'hence' comprising one tribe) despite the many dialects, multilingualisms and overlaps with surrounding and intermingled speakers of differently labelled languages.

Such images are perhaps the more entrancing with a 'speech community' that can be seen as the repository of an endangered language – an understandably value-laden topic. But does that perhaps make it all the more important for the documentation to tell it how it is - and how people use it *now* not in some notionally pure and uncontaminated past? Should the data include the diversities and contradictions, mixtures of perceptions from past and present and from differing perspectives, the invented 'traditions', unequal powers, warring viewpoints? And since it can be argued that few 'speech communities' are truly monolingual or culturally uniform (especially perhaps if their language *is* now 'endangered') will the data include the overlaps and interactions with other languages, perhaps both written and spoken, and what might once have been dismissed as 'hybrid' genres or speech? Schooled forms, popular novels, influences from 'European' genres, translations, bilingual forms, writing – all may now in practice be part of the reality of (some?) people's lives, not easily to be discounted as aberrant or 'alien'. The study of familiar western languages takes for granted that cultures, communities and languages change and interact with others. Should we demand something different, some frozen essence, once a language is classed as endangered? So too with the practices of translation, of language-switching and interpenetration, global interactions, young versus old – none of these are necessarily 'abnormal' or irrelevant. If the way 'a language' exists is in how people speak, enact, create, change and manipulate verbal resources, then the data to count might need to come from that full range, not from some idealised atemporal prior state (once again illustrating how delineating both 'speech community' and data may be inseparable from assumptions about the nature of 'language').

This also affects the question of who provides the 'data' and decides what it means. The older images of the homogeneous and unchanging 'tribe', 'language' or 'community' envisaged everyone as essentially sharing a common tradition. So, in anthropology as in language documentation, it seemed to matter little who you got your information from. The 'myths of the Bongo Bongo' (or whoever) could be elicited equally well from any member of the group. Now we are more critical. We are sensitive, hopefully, to change, manipulation, disagreements, inventions, power relations. Like others,

I have also become increasingly aware of the extent to which the processes of dictation, transcription, translation and recording are not mechanical but socially and individually shaped. [5] Here too there are always likely to be local as well as distant participants in the process with their own interests, preconceptions and variegated agendas in the formulation of data. All this is part of how language is actually used and exploited, so that documentation too is unavoidably an active, creative and far from neutral process.

One long-standing image envisaged the analysis and interpretation of data as naturally belonging to the outsider-investigators. They were the ones capable of synthesising and expertly studying the matter provided by the insider-informants. But by now many field researchers have moved away from older notions of 'informants' and 'subjects' towards acknowledging the interactive nature of research through such terminologies as 'consultants', 'collaborators', 'co-researchers'. Native scholars and thinkers analyse and organise data, and local metalinguistic conceptualisations – and, no doubt, models of language - shape how the 'data' is presented and synthesised. This complex collaborative process presumably now needs to be recognised rather than hidden. And in the world of today anyway who now is insider, who outsider?

As for how and where one finds the data, as an anthropologist I start with an inclination towards participant observation and informal interaction in addition to formal questioning, preferring that to eliciting data outside the field. But having thought further about how verbal and other cultural formulations work in practice I now recognise that as a somewhat blinkered view. For ultimately *all* these forms are humanly-produced products, and it would be misleading to privilege some as 'counting', others as not.

But equally *all* data wherever it originates has to be treated critically, with full awareness of the providers' social situatedness whether outside or within 'the field'. Looking for neutral informants channelling neutral data is unrealistic. The data we have are recorded from, assisted by, enacted by, written by, transmuted through people of particular kinds all with their own preconceptions, characteristics and agenda in terms of, for example, their age, gender, religion, education, politics and much more. And this need not exclude, where they exist, the younger non-fluent or perhaps multilingual speakers. And all are doubtless operating in the context of a developing situation of learning and changing, where the end is unlikely to be the same as the beginning – and a situation furthermore in which the investigators' and sponsors' own position, concerns and policy intentions are all part of the equation.

[5] See Finnegan 2007:chapter 10.

Not that there is anything reprehensible about this play of interests, diversities, or politics. This has always been the background to the practices of translation and of language planning, and to the struggles over what is to count as the same or separate - or original. But the controversies need to be recognised. If certain groups or forms are prioritised on the assumption, for example, that *they* are the prime bearers of 'the language'– whether the older speakers, the non-schooled, the newer literates, the bilingual children, the travelled, the stay-at-homes – then that decision needs to be manifest. Language documentation can never be, has never been, a matter of detached and objective pebble gathering, but an intensely human process of selection, analysis and, inevitably, manipulation. And it is in this context that I now find the once simple-sounding and serviceable distinction between data and metadata so much muddier that it seemed at first. As in many other areas of life there is perhaps never really 'primary' data in some sense of being 'pure', 'traditional', 'authentic'. Rather there are human beings who live in the world and formulate their interventions whether as 'speakers', 'analysts', advocates, politicians - or, more likely, a mixture of all of these and more.

8. Conclusion

The documentation of endangered languages remains an important and inspiring endeavour. But it is clearly neither a simple nor a neutral process. Like others no doubt I continue to puzzle over what can be delimited as 'language', and hence, inevitably, over what can count as 'data' and what would be needed to document a language.

Perhaps these uncertainties are unavoidable. Whatever 'language' is or is not taken to be – written text, performance, abstract system, meaning, action, people deploying resources from across the interpenetrating modes of human communication, or even, by now, an outdated term - there is clearly no single 'right' or (perhaps) cross-culturally neutral or a-political view of it. Selective choices are inevitable. But we should be clear that we *are* making them, that by going down one route we are excluding others, and – finally – that our decisions about what data counts may mean in effect tacitly lining up with some particular position about the nature and working of 'language'.

References

Austin, J. L. 1962. *How to do things with words.* Oxford: Clarendon Press.

Bakhtin, M. M. 1986. *Speech genres and other late essays.* Eng. transl. Austin: University of Texas Press.

Banti, G. and F. Giannatasio. 2004. Poetry. In Duranti 2004.

Bauman, Richard. 1977. *Verbal art as performance.* Rowley MA: Newbury House.

Bauman, Richard and Charles L. Briggs. 1990. Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19:59-88.

Bauman, Richard and Charles L. Briggs. 2003. *Voices of modernity. Language ideologies and the politics of inequality.* Cambridge: Cambridge University Press.

Bauman, Richard and Joel Sherzer. (eds) 1989. *Explorations in the ethnography of speaking*. 2nd edn. London: Cambridge University Press.

Dalby, David. 1999/2000. The Linguasphere Register of the World's Languages and Speech Communities. 2 vols. Hebron, Wales: Linguasphere Press.

Duranti, Alessandro. (ed.) 2004. *A companion to linguistic anthropology*. Oxford: Blackwell.

Feld, Steven and Aaron A. Fox. 1994. Music and language. *Annual Review of Anthropology* 23:25-53.

Finnegan, Ruth. 1967. *Limba stories and story-telling.* Clarendon Press: Oxford (also Greenwood Press: Connecticut, 1981).

Finnegan, Ruth. 1970. *Oral literature in Africa*. Clarendon Press: Oxford (also Nairobi: Oxford University Press, 1976).

Finnegan, Ruth. 1977. *Oral poetry: its nature, significance and social context*. Cambridge: Cambridge University Press. [2nd edition Bloomington: Indiana University Press, 1992.]

Finnegan, Ruth 1988. Literacy and orality: studies in the technology of communication. Oxford: Blackwell.

Finnegan, Ruth. 1989. *The hidden musicians: music-making in an English town*. Cambridge: Cambridge University Press. [2nd edition, Middletown: Wesleyan University Press, 2007.]

Finnegan, Ruth. 1992. Oral traditions and the verbal arts: a guide to research practices. London: Routledge.

Finnegan, Ruth. 1998. *Tales of the city: a study of narrative and urban life*. Cambridge: Cambridge University Press.

Finnegan, Ruth 2002. Communicating: the multiple modes of human interconnection. London: Routledge.

Finnegan, Ruth. 2006. Which comes first: the words, the music or the performance? Conference paper Rio de Janeiro, May 2006 [in Portuguese translation in Cláudia Neiva de Matos, Elizabeth Travassos, and Fernanda Teixeira de Medeiros (eds) forthcoming. *Palavra cantada.*]

Finnegan, Ruth. 2007. *The oral and beyond: doing things with words in Africa.* Oxford James Currey/ Chicago: University of Chicago Press.

Foley, William. 2003. Genre, register and language documentation in literate and preliterate communities. *Language documentation and description* 1:00-00

Franchetto, Bruna. 2006. Ethnography in language documentation. In Gippert et al. 2006:00-00.

Gippert, Jost, Nikolaus P. Himmelmann and Ulrike Mosel. (eds) 2006. *Essentials of language documentation*. Berlin: Mouton de Gruyter.

Hanks, William F. 1996. *Language and communicative practices*. Boulder: Westview Press.

Harris, Roy. 1987. *The language machine*. London: Duckworth.

Harris, Roy and G. Wolf. 1998. *Integrational linguistics. A first reader*. Oxford: Pergamon.

Haviland, John B. 2004. Gesture. In Duranti 2004:00-00.

Hill, Jane H. 2006. The ethnography of language and language documentation. In Gippert et al. 2006:00-00.

Himmelmann, Nikolaus P. 2006. Prosody in language documentation. In Gippert et al.. 2006:00-00.

Hodge, Robert and Gunther Kress. 1993. *Language as ideology*. 2nd edition. London: Routledge.

Hymes, Dell. 1977. Foundations in sociolinguistics: an ethnographic approach. London: Tavistock.

Kendon, Adam. 2004. *Gesture: visible action as utterance.* Cambridge: Cambridge University Press.

List, George. 1963. The boundaries of speech and song. *Ethnomusicology* 7, 1:1-16.

McNeil, David (ed.) 2000. *Language and gesture.* Cambridge: Cambridge University Press.

Tracey, K. (ed.) 1999. Language and social interaction at the century's turn. Special issue. *Research on Language and Social Interaction* 32, 1/2.

Treitler, Leo. 2003. With voice and pen. Coming to know medieval song and how it was made. Oxford: Oxford University Press.

.