elpublishing

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 7*. Editor: Peter K. Austin

Documenting sign languages

ADAM SCHEMBRI

Cite this article: Adam Schembri (2010). Documenting sign languages. In Peter K. Austin (ed.) *Language Documentation and Description, vol 7*. London: SOAS. pp. 105-143

Link to this article: http://www.elpublishing.org/PID/083

This electronic version first published: July 2014

Documenting sign languages

Adam Schembri

1. Overview

This chapter is an introduction to documentary sign linguistics. I am going to focus mostly on the British Sign Language Corpus Project, based at the Deafness, Cognition and Language Research Centre, University College London — in particular, issues to do with data collection (e.g., the Observer's Paradox) and the notion of having an open access archive[1].

As background, I have worked primarily on Australian Sign Language (Auslan); it has been easy for me to move between British Sign Language (BSL) and Auslan because they are arguably dialects of the same sign language. New Zealand Sign Language (NZSL) also forms part of the BSL family, so the three sign languages might best be considered dialects of the same variety (Johnston 2003). They are quite different from American Sign Language (ASL); BSL and ASL have quite different histories and are not completely mutually intelligible.

Unlike some other corpus projects (e.g., the Auslan Archive[2] or the Sign Language of the Netherlands (NGT) Corpus[3]) which received funding specifically for the creation of a corpus, the funding for the BSL Corpus Project comes from a grant by the Economic and Social Research Council

[1] Thanks to the following researchers whose work influenced the BSL Corpus Project research design: Trevor Johnston, Ceil Lucas, Onno Crasborn, and David McKee. I am also grateful to the BSL Corpus Project co-investigators and the Deaf Community Advisory Group members whose input has been invaluable: Kearsy Cormier, Margaret Deuchar, Frances Elton, Donall O'Baoill, Rachel Sutton-Spence, Graham Turner, Bencie Woll; Linda Day, Clark Denmark, Helen Foulkes, Melinda Napier, Tessa Padden, Gary Quinn, Kate Rowley, Lorna Allsop. This paper would not have been possible without the work of the BSL Corpus Project team: Jordan Fenlon, Ramas Rentelis, Rosemary Stamp, Sally Reynolds, Jenny Wilkins, Jacqueline Parker, Carolyn Nabarro, Mark Nelson, Mischa Cooke, Melinda Napier, Jeff Brattan-Wilson, Avril Hepner, Evelyn McFarland, Dawn Marshall, Sarah Lawrence and Breish Rowe. I am grateful to Robert Adam and Trevor Johnston for comments. Thanks lastly to all the members of the British Deaf community who agreed to participate in the project.

[2] http://www.auslan.org.au/about/corpus/

[3] http://www.ru.nl/corpusngtuk/

Alan Schembri 2010. Documenting sign languages. In Peter K. Austin (ed.) *Language Documentation and Description, Vol 7*, 105-143. London: SOAS.

(ESRC) as part of a research project to create a corpus while conducting linguistic research into BSL. The key research questions focus primarily on issues related to sociolinguistic variation and change in BSL. Obviously corpus data lends itself rather well to this kind of study, and I will explain below why we have such an interest in variation.

2. Sign languages: myths and misunderstandings

To begin, let us deal with some myths and misunderstandings about sign languages. First of all, sign language is not universal. There are different sign languages in use all over the world. There actually **is** something called International Sign, which is a pidgin that has developed as a result of regular language contact, particularly between users of European sign languages. It is only very partly documented in the form of a dictionary produced by the World Federation of the Deaf (WFD) (Rubino et al. 1975), and it is used in WFD meetings when Deaf representatives from different parts of the world come together. Thus, there is a kind of international sign language, but it is a natural pidgin that has developed from language contact. ASL is also quite often used as a sign language lingua franca, particularly in academic circles.

As noted above, ASL and BSL are not mutually intelligible, although it has to be said that, because of high degree of iconicity in sign languages (i.e., a lot of signs have a close relationship between their meaning and their form), a significant number of signs are shared across different sign languages (see, for example, Guerra Currie et al. 2002), and Deaf people are able to negotiate a common vocabulary and communicate basic information rather quickly upon first meeting, particularly in a European context. But that does not mean that sign languages as used normally within their different respective Deaf communities are mutually intelligible. I have a basic level of survival ASL but, before studying it, the language would not have been very intelligible to me as a user of Auslan/BSL.

Secondly, sign languages are not codes for spoken languages. Often people say to me: 'it is amazing – when you watch the news on the BBC, you see the sign language interpreter and they seem able to keep up with the really rapid spoken English of the newsreader'. That is because they are interpreting; they are not actually representing in signed form every English word that the speaker is using. Sign languages have different vocabularies and different grammars from the spoken languages in the surrounding community. That does not mean, however, that there is no influence from the spoken language of the surrounding hearing community on Deaf community sign languages. In particular, BSL is often accompanied by silent mouthing of equivalent English lexical items, while signing is occurring. In fact, there is a considerable

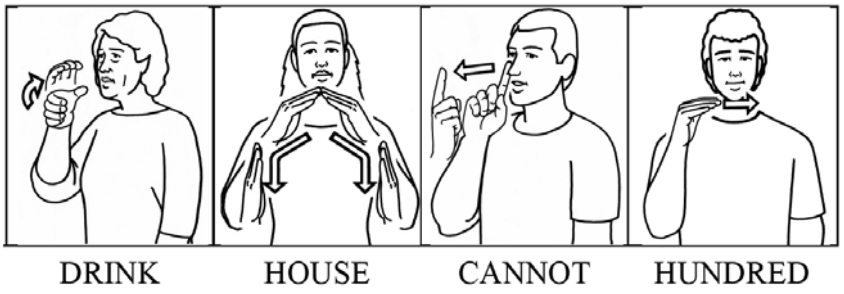
influence from spoken languages on sign languages, but the grammar and the vocabulary does not line up morpheme-by-morpheme.

There **are** sign language codes for spoken languages that have been created by educators of Deaf children, for example. Thus, in Australia, there is the Australasian Signed English system; this is a contrived, artificial sign system that was created by educators to represent English grammar, morpheme-by-morpheme (Jeanes & Reynolds 1982; Johnston & Schembri 2007). But that is not the natural language of the Australian Deaf community, which is Auslan. Australasian Signed English has itself influenced Auslan, but it is not something that is any community's first language. It is a manual code for spoken English.

Thirdly, sign languages are quite iconic (as mentioned earlier), but that does not mean that they are just a sequence of pantomime or gestural exchanges. Obviously if they consisted of nothing apart from very iconic gestures, then there would not be difficulties communicating across different sign languages, or between non-signers and signers. This is not the case.

Obviously sign languages are much more iconic than spoken languages – many signs have some kind of relationship between form and meaning. For example, the BSL signs HOUSE, EAT, DRINK and TIME are easily recognisable (DRINK and HOUSE are shown below), and have the same form in ASL and a number of other sign languages. So there **are** a lot of iconic signs, but there are also signs that are completely arbitrary – and this is what gives rise to the mutual unintelligibility. So you probably would not be able to guess the meaning of the BSL signs for CANNOT, BROTHER, HUNDRED and ENGLAND (CANNOT and HUNDRED are shown below). These signs are also quite different, quite unrelated, in ASL.

Figure 1: Iconic and non-iconic BSL signs



Fourthly, sign languages were not invented by hearing people. You would be surprised how often sign language interpreters get asked, 'So do you know Braille as well?' One difference between sign languages and Braille is that Braille was invented by Louis Braille, a Frenchman who created a system for representing written language in raised dots for blind people to read. Sign languages were not invented (and are not codes for spoken languages, as explained above). Aside from the artificial sign systems which educators have created, sign languages – like spoken languages – have developed naturally when Deaf people have come together to form a community. This explains why they are not universal, and why, for example, ASL and BSL are different – because nobody planned them; they developed naturally and independently within their respective Deaf communities.

In particular, sign languages developed when Deaf children were brought together for the first time in Deaf schools. In the European context, this started in the 1760s in both Scotland and in France (Kyle & Woll 1985). The French opened a public school around this time, while the Scottish school was a private academy for the children of the wealthy. Prior to industrialisation and urbanisation in Europe, most Deaf children (unless they happened to encounter each other or were from the same family) would have been fairly isolated, because deafness is a fairly low-incidence phenomenon. Most Deaf children would have grown up as isolates in rural settings, in villages, etc.

During the industrialisation and urbanisation that happened particularly in the UK and other parts of Europe in the 18th century, more and more people came together in cities, and indeed Deaf people – because they would meet each other – began to form communities. Prior to the establishment of the first school in France, we know that there was already a signing community in Paris, for example, because the first teachers who taught in the school learnt the sign language from the Deaf community (Lane 1984). That is, Paris was a large enough urban centre for Deaf people to have formed a community naturally, and for a sign language to have evolved. This also seems to have occurred in the UK: there are records of sign language use prior to the first schools (Kyle & Woll 1985), but it really was the schools that brought together sufficient numbers of Deaf children for the first time (Woll & Ladd 2003).

Schools were also centralised places where transmission of languages could occur, because (as discussed below), sign languages have perhaps always been endangered languages. Even in Western communities, for example, no more than around 5-10% of the adult Deaf community themselves have Deaf signing parents (Mitchell & Karchmer 2004). This means a majority of Deaf people (90% or more) acquire sign language not vertically down the generations but horizontally from peers. Traditionally,

most Deaf people have only come into contact with a sign language when they enter a centralised Deaf school and meet other Deaf children.

Often the Deaf people from Deaf families (there are a small number of multi-generational Deaf families) are like the carriers of the language – they transmit it down the generations, through the centralised schools. Below I will discuss why that does not always happen, and why it is happening a lot less often now, because of de-institutionalisation of Deaf children and the mainstreaming of Deaf children's education in particular. Deaf children are often sent to local schools with a sign language interpreter – in fact, in the UK, such 'interpreters' do not need to be qualified but can just be someone with a relatively basic level of skill. Today, it is often those people who serve as the language models for the Deaf children, rather than other Deaf signers.

Oddly enough, despite the current fad for 'baby signing' among hearing parents with hearing children (e.g. http://www.signingbaby.com/main/), some Deaf children are denied access to sign language because of a mistaken belief that it will affect their speech development. In fact, Deaf children from Deaf signing families tend to outperform other Deaf children on tests of English literacy (see, for example, Hoffmeister 2000). Meanwhile middle class parents with hearing children are hot-housing their children in ASL and other sign languages in the belief (and there is no evidence to support this) that it will actually kick-start their language development earlier than is normal and/or lead to significant improvement in their cognitive development generally.

This belief appears to be based on data from the 1980s that seemed to suggest that the first signs in children acquiring a sign language appeared a couple of months before the first words of children acquiring a spoken language (e.g., Folven et al. 1984). The first word generally appears between 12 to 18 months of age. There were reports in the early 1980s that signing children were producing their first signs at the age of 9 or 10 months, considerably earlier.

Those research findings are now hotly debated. In fact, it may be that the criteria for defining the first signed lexical item were not being as strictly applied as they needed to be, and that actually first signs appear around the same time as first words **(**e.g., Petitto 2000). Most of the available research suggests that children learning sign languages appear to progress through the same kinds of milestones as children learning spoken languages. As far as the young brain in concerned, language is language – it does not appear to matter whether it is signed or spoken.

Although research on this issue is ongoing, this also appears to be true for the adult brain. Considerable research by neuroscientists (including those at the Deafness Cognition and Language Research Centre) who carry out scanning studies using functional magnetic resonance imaging of adult Deaf

people's brains indicates that, while sign languages recruit some extra parts of the right hemisphere of the brain, the main processing of sign language seems to occur in the same parts of the left hemisphere as it does for spoken language (Emmorey 2002).

Deaf signing people may experience brain damage in the anterior areas of the language regions of the left brain hemisphere as a result of stroke, and this usually results in Broca's aphasia (i.e. grammatical omissions in their sign language, and difficulties in sign articulation), in the just the same way that hearing-speaking people get Broca's aphasia as a result of damage to that region. In other words, key parts of the left hemisphere are important for both spoken and sign languages (Emmorey 2002). That shows again that, for the brain, language is language – it does not seem to matter whether it is in the visual-gestural modality or the spoken language modality.

3. Our focus here: Western urban sign languages

BSL is an example of a natural sign language from a 'macro-community' (a large national Deaf community) – in the UK there are BSL users in Northern Ireland, Scotland, Wales and England. It is an identifiable community (although there is considerable regional variation, mostly lexical in nature), because there is a sense in the community that BSL is one language with many dialects. The British Deaf community is nation-wide with a national organisation called the British Deaf Association, which holds regular conferences and national meetings. There is thus a real sense of national British Deaf identity, and BSL is the language of a macro-community.

We do not actually know how many Deaf people use BSL. Estimates are often based on the assumption that there are about 1 in 1,000-2,000 people in the community who use a sign language, which would give 30,000-60,000 Deaf BSL users. No one actually knows, however, how many BSL users there really are. Some recent research in Australia suggests that the estimate of the number of signing Deaf people should be reduced to at most 1 in 3,000 of the whole population (Johnston 2004). If that were the case in the UK, there would be about 20,000 Deaf BSL users across the country; we simply do not know.

This is in contrast to sign languages of 'micro-communities'. Language documentation projects have started in these smaller sign language communities, often consisting of no more than a single village, as well. The Al-Sayyid Bedouin Sign Language (ABSL) is quite a well-known case. Less known are Adamarobe Sign Language in Ghana, Kata Kolok Sign Language from Bali, and so on. These are small communities in which, due to

intermarriage, deafness has become disproportionately common in the population.

In the ABSL case, for example, these are Bedouin people in the Negev desert in Israel, where – due to traditions of marrying cousins and other relatives – deafness has an incidence of around 10% of the population (Woll & Ladd 2003). Every family has someone who is Deaf, and as a consequence many hearing people use sign language to varying degrees.

Nicaraguan Sign Language (Idioma de Señas de Nicaragua, or ISN) is a different case: this is an emerging sign language. The processes that happened in the UK in the 18th century have been happening in Nicaragua since 1979 and have been well documented (Kegl et al. 1999). In 1979, the Sandinista government opened up the first school for Deaf children; they were brought together for the first time and a pidgin sign language developed. Then, as more and more younger children came into the programme, they were exposed to the pidgin and began to creolise the sign input into a more systematic variety of sign language called ISN.

The Nicaraguan case has been well documented, and even Steven Pinker talks about it in his best-selling book *The Language Instinct* (Pinker 1994). It is perhaps the best documented case of a micro-community sign language: it is believed, for example, that ABSL is only about three generations deep – the first deafness appeared about 70 years ago, but no-one was there to document what happened when the first Deaf children were born into the community. For ISN, researchers have been working in Nicaragua since the 1980s, not long after the community there first came together.

Note that I am also not focussing here on artificial sign systems (e.g. Signing Exact English, Australasian Signed English); I mentioned them briefly earlier. There are also natural outcomes of contact between sign languages and/or spoken languages**,** e.g. the International Sign pidgin. Also, within the UK there is a variety of signing called Sign Supported English (SSE), in which a person would speak in English while supporting his/her English with signs (a kind of mixed system which has arisen naturally as a result of contact between the two languages).

Finally, there are alternate sign languages such as Warlpiri Sign Language from Central Australia – these are also not the sign languages I work on. In the Warlpiri community, female members undergo a speech taboo of up to six months (or even a year or two) during which they are not allowed to speak. Women in the women's camp often use a sign language, which is similar to a signed version of Warlpiri (Kendon 1988). Warlpiri Sign Language has been well documented and it seems to be quite similar in complexity to the sign languages of Deaf communities. Alternate sign languages, however, arise in quite different sociolinguistic situations, and are used only by hearing people

(monastic sign languages used traditionally by monks sworn to a vow of silence are another example of a group of alternate sign languages).

In a way, it might be argued that the sign languages of micro-communities (Al-Sayyid Bedouin, etc.) are a blend of Deaf community sign languages and alternate sign languages, because in these micro-communities there are many hearing users who use signed varieties showing a range of language contact characteristics. It is also possible that the different sociolinguistic situations have various impacts on the grammar of those sign languages, but they have not been very well documented thus far, so we are still finding out about them.

4. Defining 'documentary sign linguistics'

Woodbury (2003) notes that documentary linguistics of spoken languages emerged since the late 1990s, in tandem with the following changes:

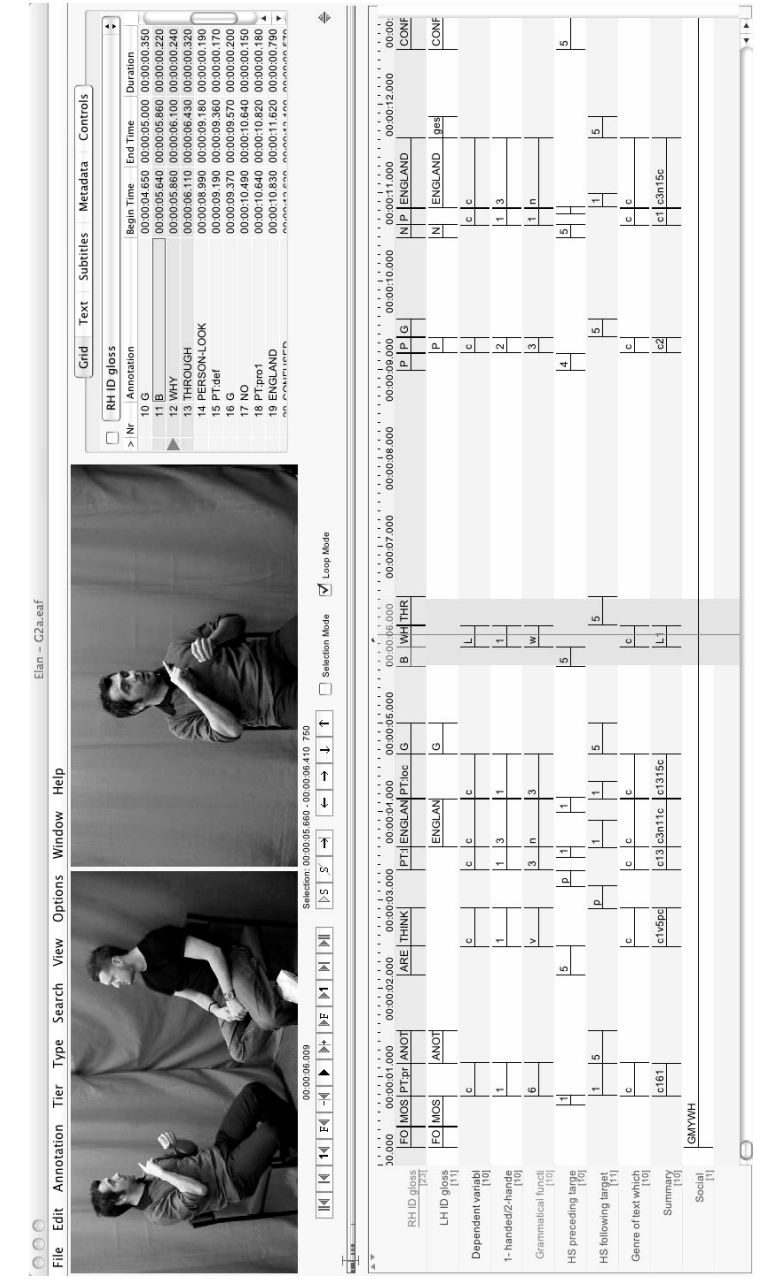
1. Major changes in the technology of data representation and maintenance;
2. New attention to linguistic diversity;
3. Concern about language endangerment;
4. Growing awareness of the needs of stakeholders outside the academic community.

Each of these is also true of sign language research. I searched for the term 'documentary sign linguistics' on-line and I did not find anything ('sign language documentation' is also relatively rare), so this paper may be the first time the term has actually been used. In fact, there is increasing interest in corpus-based approaches to sign languages within sign linguistics, but there has not been awareness of re-thinking these approaches as examples of 'documentary sign linguistics', having much in common with language documentation projects for minority and endangered spoken languages. I now consider each of Woodbury's points in turn.

4.1 For sign languages, software such as ELAN has made documentary sign linguistics possible

Changes in data representation for sign languages have lagged a little behind those in spoken language linguistics, as we have had to wait for the wider use of digital video and the development of related computer technology to make it possible to collect and store large amounts of video data. Since the introduction of ELAN (Eudico Linguistic Annotator) and related types of video annotation software, developments have been much more rapid.

Figure 2: ELAN annotations



Originally developed by the Max Planck Institute for Psycholinguistics in the Netherlands as a way of recording spoken language and co-speech gesture, ELAN has really revolutionised sign language linguistics.

In Figure 2, there is an example of an ELAN annotation file from the BSL Corpus Project. This now enables sign language researchers to share with each other both our analyses of the data and the data itself in digital video form. For sign languages, documentation projects have only really become possible with the advent of this type of technology.

There is other sign language annotation software, such as SignStream[4], and iLex[5], which combines a video annotation tool with database facilities. However, ELAN is perhaps the most widely used at the moment, and is rapidly becoming a standard tool within sign language research.

4.2 Just in time!

ELAN has come along just in time, because sign language researchers have not been operating effectively for quite some time! The phonological structure of sign languages was first analysed in 1960, when William Stokoe published a monograph on ASL (Stokoe 1960) and showed that the signs of ASL could be broken down into sub-lexical units – handshape, location and movement – and that each lexical sign was a combination or re-combination of a limited set of handshapes, locations and movements (like the phonological features of spoken languages).

That discovery took a while to catch on. Sign language research began to develop further in the USA in the early 1970s, led by a team based in the laboratory of Edward Klima and Ursula Bellugi at the Salk Institute for Biological Studies (Klima & Bellugi 1979). It began in Europe in the late 1970s, with research teams in Bristol and Edinburgh in the case of the UK, for example. This means it is really a relatively new area, and only in the last three decades has sign language research taken off. In the last decade, sign language research has begun in many countries in the developing world where previously there was little research.

However, there has not been sufficient time nor apparent desire for any common written form of sign language to be adopted by sign language researchers: there is no sign language equivalent of the International Phonetic

[4] http://www.bu.edu/asllrp/SignStream/

[5] www.sign-lang.uni-hamburg.de/lrec2008/pdf/lrec2008_hanke.pdf

Alphabet. There have been notation systems developed, such as the Hamburg Notation System shown in Figure 3 for the BSL sign CENTRE:

Figure 3: Hamburg Notation System

In this example, the line underneath the illustration represents the sign's formational features in a Hamburg Notation System (HamNoSys) representation, created by the Institute for German Sign Language in Hamburg; it attempts to represent the handshapes, locations and movements of individual signs with iconically motivated symbols.

For example, in the notation of the BSL sign CENTRE, the first symbol of the representation rather iconically represents the signer's left handshape showing a flat hand configuration with the thumb extended; the second symbol represents the direction of the fingers (the black line represents the body and the arrow means pointing away rightwards from the body); the third symbol shows the palm orientation is upwards, etc.

Unfortunately, although HamNoSys has existed for around 20 years, it has not been universally adopted by sign language researchers. It is not clear why this is the case, but it may partly reflect the fact that there are other sign language writing systems that have been proposed as well, including Stokoe Notation[6], and Sutton SignWriting[7] created by Valerie Sutton from her earlier work on DanceWriting.

[6] http://en.wikipedia.org/wiki/Stokoe_notation

[7] http://en.wikipedia.org/wiki/SignWriting

However, apart from specifically phonetic or phonological analyses of a sign language, or to record phonological variation, we have effectively leap-frogged the need for a written representation based on formational characteristics for sign language documentation, because now we can actually see the data itself using software such as ELAN (Johnston 2010).

In the sign language linguistics literature (such as Johnston & Schembri 2007), signs are represented by means of glosses in upper case English letters, with a few additional symbols. This is sufficient for studying the syntax of sign languages, but obviously it is a very poor way to represent them. However, that is the kind of system that has become widely used, and it has meant that sign language researchers have very limited access to formational information about each other's data.

Now, with digital video and video annotation software such as ELAN, it is becoming more widely expected that researcher will be able to present their primary data.

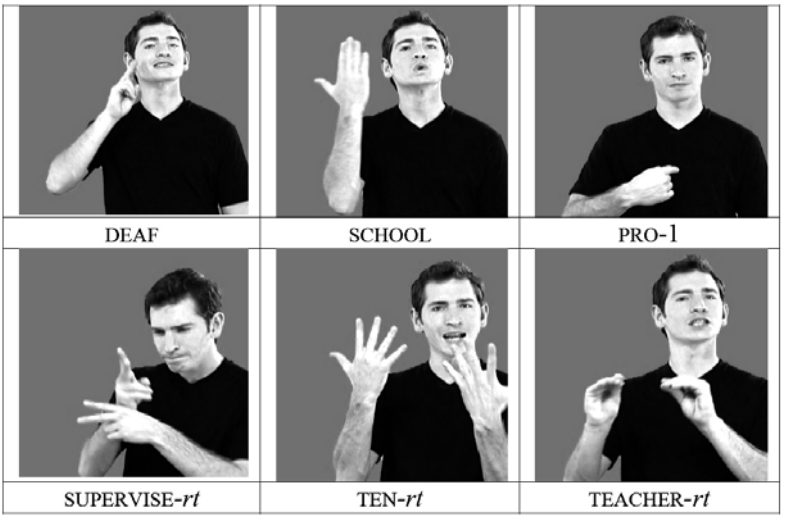
4.3 Sign languages as a test case for language diversity

Linguists often look towards sign languages as an interesting test case for ideas about linguistic universals (e.g. Pinker 1994), although perhaps not as often as they should (see Evans & Levinson 2009). Currently, there appears to be what I consider a premature consensus in some parts of the wider linguistics literature that many of the same phenomena found in spoken languages are also found in sign languages (e.g., Fromkin et al. 2006). Like Evans & Levinson (2009), I think the problem here is that some of the generalisations across signed and spoken languages require a dangerous degree of abstraction away from the data itself. Let's look now at analyses proposing that sign languages have verb agreement systems, for example.

I say that such claims are premature, because actually there is not a lot of consensus within the sign language linguistics field itself about how to analyse the structures that have come to be known as agreeing verbs (cf., Sandler & Lillo-Martin 2006; Liddell 2003). For example, the verb SUPERVISE in BSL is considered in mainstream analyses (particularly those conducted by generative and formal syntacticians) to be an example of an agreement verb (e.g., Padden 1988; Sutton-Spence & Woll 1999), marking for person agreement. In the functional/cognitive tradition which is now emerging (Liddell 2003; de Beuzeville et al. 2009), however, it has been proposed that what is going on with verb signs like SUPERVISE in BSL is typologically different from the agreement seen in spoken languages – BSL and other sign languages have a system that is unique in its specific details to sign languages (although similar combinations of gesture and language occur in speech).

Let's look at an example of the verb SUPERVISE in action, as in Figure 4. The example means: 'At the Deaf school, I supervise 10 teachers'. Look closely at the sign SUPERVISE. You can see that from that point onwards in the discourse, the signer has shifted his signing to the right. During SUPERVISE, he directs the sign clearly towards the right (this is represented here by the abbreviation *–rt* attached to the gloss). In its unmodified or citation form, the sign is produced directly in front of the signer's chest, with the fingers pointing forward, but this is not what we see here. The signer's eye gaze is directed towards the right as well, and there is a change in facial expression during this sign. You will notice too that, although his facial expression changes back to neutral and eye gaze returns to the addressee, he actually signs TEN and TEACHER slightly towards the right.

Figure 4: 'Agreement' in BSL



Some linguists argue that this is an example of an agreement system. In this case, the verb sign would be showing object agreement: the sign SUPERVISE is directed towards a location associated with the object noun phrase meaning 'ten teachers'.

Corbett's (2006) work on the typology of agreement systems, however, suggests that there has to be some formal or semantic property of the controller noun phrase that is realised on the target verb. While one might want to argue that it is a formal property of the signed noun phrase TEN TEACHER that it is produced on the right, it actually is not necessarily a formal property of this noun phrase: it just happens that, in that instance, the

signs are moved to the right, perhaps as a result of assimilation following the sign SUPERVISE. It would be perfectly acceptable to sign the same sequence with the noun phrase TEN TEACHER produced directly in front of the signer.

The sign SUPERVISE can also be directed at the location of referents that are present in the room around the signer. I could say that I am supervising Oliver (the interpreter here to my left) and I could direct the sign SUPERVISE to his actual present location to produce the sequence PRO-1 SUPERVISE-*lf* O-L-I-V-E-R. But Oliver's real world location next to me on my left at this very point in time is not a semantic or formal property of the controller noun phrase O-L-I-V-E-R (I have used the British manual alphabet to spell out his name, because I don't know his name sign): it is a characteristic of the real world referent of that noun phrase.

I would not want to argue that the locations of referents referred to by noun phrases in sign languages are part of the grammar of sign languages – otherwise the grammar would be completely unbounded, because it would involve the locations of every possible referent that could be talked about. What I would like to argue, and what some sign language linguists from a cognitive and functional perspective argue (supported by the spoken language evidence provided by typologists such as Corbett), is that actually what we have here is a typologically unique fusion of a linguistic element with a pointing gesture in the same lexical item. This system is used within sign language as a referent-tracking device, like agreement systems arguably are too in spoken language, so you can identify referents, but it actually involves a combination of a categorical lexical sign with a gradient pointing gesture. In sign languages signs are made with the hands and you can move your hands around in space, you can point them at locations associated with referents in a way that you cannot do with spoken languages. I cannot, for example, direct the word *supervise* towards Oliver's location in saying "I supervise Oliver" – sound travels in all directions, so it is not physically possible to direct the sound specifically at his location.

Sign languages probably do this because it is physically possible in the visual-gestural modality. After all, it is a great system for referent tracking – a very clear visual system that uses space and the location of real world referents. It is not possible for spoken language lexical items to do this – sign languages do it because they can (and speakers use co-speech gestures when they can too). Sign language lexical items, I like to think, are iconic and indexical in this way because they can be. So, unlike what many introductory textbooks of linguistics say (e.g., Fromkin et al. 2006), it is not actually a defining feature of languages that they should be arbitrary – it is just that it is very difficult to be iconic in spoken languages because most of the objects we want to refer to do not make distinctive sounds that speakers

can imitate. But many of them have visual characteristics that I can imitate: with PROJECTOR (to refer to an object in this room), there is an aspect of the form of the sign that represents what is actually happening: the hand opens to represent the movement of the light being projected onto the screen.

I would argue that getting it right about sign language grammars is very important for understanding linguistic universals. The system of *indicating verbs* (a term proposed by Liddell 2003, to replace *agreeing verbs*) is, I think, unique to sign languages and represents a real modality-related typological difference. I do not think it sits well in a classification as an agreement system. I think we lose something about understanding the diversity of human languages if we treat it that way. And indeed, to come back to the point about documenting sign languages, an improved understanding of indicating verbs is emerging as a result of sign language documentation projects, such as the Auslan Archive (de Beuzeville et al. 2009).

Within the history of sign linguistics itself, there really was a very strong focus in the first two or three decades on Western urban sign languages, on sign languages of macro-communities. Most of the research happened in North America, Europe and Australasia, and there been some premature claims about what might be universal to sign languages based on this research on macro-community sign languages. For example, all the Western sign languages that have been described so far do appear to have indicating verb systems, but it has been claimed that ABSL does not (Meir et al. 2007), although this may be because the language is a pidgin that has not yet developed this system. Nevertheless, there are other micro-community sign languages that have not been documented or described sufficiently, so in order for sign language typology to advance we need to understand more about these other types of sign languages. And the best way to do that is to document these sign languages as thoroughly as possible.

4.5 Sign languages: always and increasingly endangered?

There is a growing awareness that many sign languages are endangered (Johnston 2004; Nonaka 2004). In a way, one could argue that all sign languages have always been endangered, due to the interrupted patterns of transmission that have occurred with most macro-community sign languages, and the active attempts to suppress their use in Deaf education during much of the latter half of the 19th century until the late 20th century (Woll & Ladd 2003). There has also been an argument, perhaps, that each generation re-creolises sign language (Fischer 1978), because in each generation there is a majority of non-native signer input. So there has been a

suggestion that, in a sense, sign languages – because of the nature of their transmission – are trapped in this constantly re-creolising cycle.

4.5.1 The suppression of sign languages

In one way, sign language communities are fragile: transmission can be interrupted relatively easily. In New Zealand, for example, prior to the establishment of the first school for the Deaf in 1880, it is believed that there were Deaf people (either Deaf immigrants from the UK or New Zealand-born Deaf people) who used a variety of BSL and were instructed by a British tutor (Collins-Ahlgren 1989). We believe that some variety of BSL-based signing was being transmitted to these early generations of Deaf people in New Zealand prior to the establishment of the first school. It would have represented an ideal opportunity for all of the Deaf children around New Zealand to come together, as happened in Nicaragua in the late 20th century. Deaf schools at the time were residential, which is why they were so important for the creation of Deaf communities, because Deaf children came together and lived on-site for most of the year.

In 1880, however, in Milan an international congress of educators of Deaf children (almost all of whom were hearing) decided that sign language as a means of instruction was not the best method for Deaf education, and that there should be a focus on spoken language development for Deaf children, and that Deaf children should learn to lip-read and to speak as much as possible. The influence of this recommendation was felt far and wide, and began a global movement to ban the use of sign languages in classrooms with Deaf children (Lane 1984).

Prior to the late 18th century and early 19th century, sign languages were used as a means of instruction in many Deaf schools; this began to be phased out from around 1880. In fact, at the first school in New Zealand in Christchurch (now known as the Van Asch Deaf Education Centre), the teacher Gerrit van Asch was a strong oralist (he believed only in education that focussed on the development of speech and speech-reading skills). He forbade the use of sign language in the Deaf school and, not only that, he knew about the transmission of sign language from peer to peer, so he banned any Deaf child who knew any sign language from being accepted into the school (Collins-Ahlgren 1989).

The school records for the Royal Institute for Deaf and Blind Children and the Victorian College for the Deaf in Australia indicate that some New Zealand Deaf children began arriving in Australian Deaf schools in Melbourne and Sydney from about 1880, because they were not allowed to

go to the New Zealand Deaf school. That ban on sign language use in Deaf education in New Zealand continued from 1880 to 1979.

NZSL is a very interesting case because some of the lexicon and derivational morphology that occurs in both Auslan and BSL was lost in NZSL. For example, Auslan and BSL show numeral incorporation: number signs such as TWO and THREE may be (optionally) incorporated into the signs WEEK or YESTERDAY to create different signs for lengths of time, such as IN-TWO-WEEKS'-TIME or THREE-DAYS-AGO. As I learned firsthand when teaching a course on sign language linguistics in Auckland for the New Zealand Sign Language Tutors Association, these usages are entirely missing from older generation NZSL users who sign instead THREE DAY AGO as three separate lexical items and do not combine the number handshape into the sign for YESTERDAY in this way (anecdotal reports suggest, however, that this is less true of younger NZSL signers who now sometimes use forms with numeral incorporation – Rachel McKee, personal communication). This is one impact of the break in transmission: an aspect of the morphology in the BANZSL family was not traditionally present in the sign language used in New Zealand. Unique signs, such as the number signs NINE and ELEVEN, also occur in New Zealand – these are different from signs in any part of the BANZSL family, apparently independently created by the Deaf children at the Van Asch school who had no BSL language model in the classroom or dormitories.

The situation with NZSL, which is not unique, is an excellent example of what happened in many schools across the Western world where sign language use was banned. Thus, for much of their history, even sign languages of macro-communities have been suppressed and endangered.

4.5.2 Deaf communities as endangered communities?

Sign languages are now endangered in a new way as well. I find this upsetting and difficult to write about, but in Western countries the demographics of deafness are changing. The number of Deaf children being born is dropping quite significantly. Thus, Deaf communities in Western countries are endangered, not just the sign languages they use.

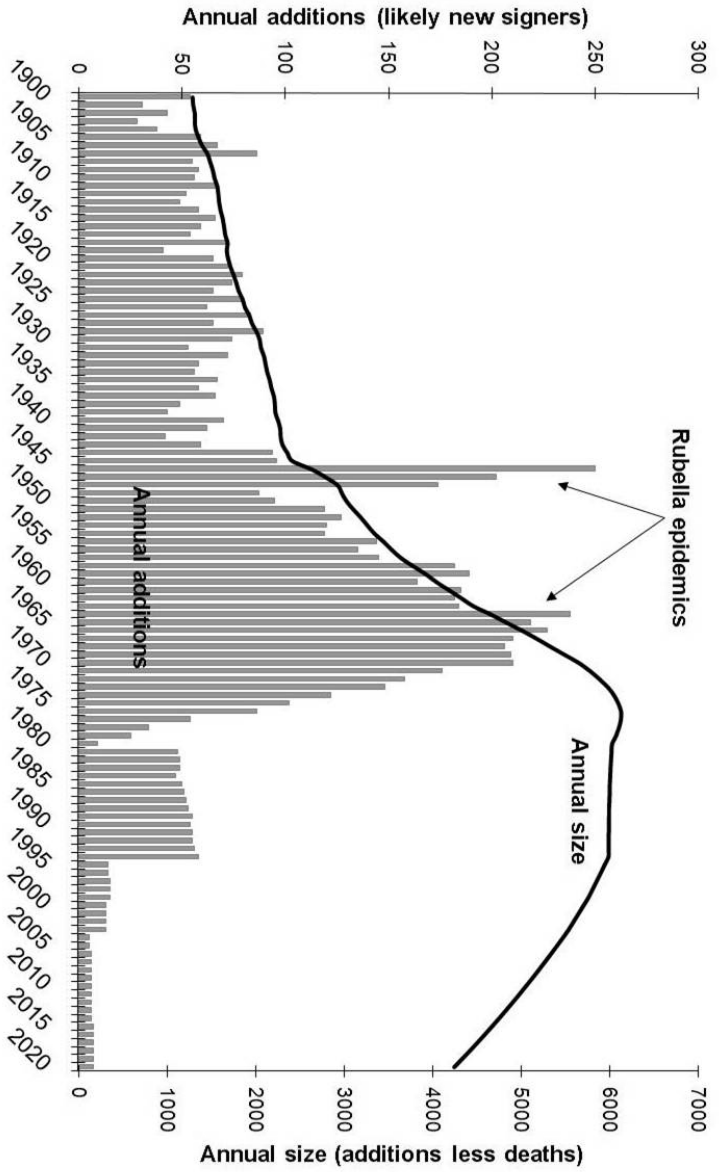
As noted above, about 5% of the adult Deaf community come from Deaf families, so these are genetic causes of deafness that are passed down from one generation to another. Genetically-caused deafness can also occur spontaneously: I have a Deaf friend in Australia whose sister is also Deaf. Although no-one else in their family was Deaf before them, they both appear to carry a gene for deafness that results in both of them being Deaf. They have been informed by genetic counsellors that they are highly likely

to have Deaf children themselves as a result. Much of the adult Deaf community today in countries like the UK, Australia and the US are Deaf, however, as this is a result of other factors – in particular, maternal rubella, a very common cause of deafness.

As can be seen from Figure 5, the last maternal rubella epidemics in Australia were in the 1960s and 1970s (Johnston 2004). Maternal rubella, before procedures such as MMR (measles, mumps and rubella) vaccination, resulted in significant numbers of children being born deaf and so there are 'deaf baby booms' in the population as a result. Similar patterns of rubella epidemics occurred during this period in other Western countries. But maternal rubella has been pretty much eradicated by advances in modern medicine.

Also, improving hearing aid technology and the use of cochlear implants has led to a new movement back towards a focus on speech and listening skills in Deaf education, often to the exclusion of sign languages. I would argue that, even when children are given a cochlear implant, they have a right to access a sign language, partly because the success rate of cochlear implants for speech development is highly variable (Marschark & Spencer 2003). There is no guarantee that a child is going to learn speech, and I believe that they should be raised bilingually anyway – they should be exposed to a sign language as well, because (among many other reasons) as adults they may choose to use the sign language as their primary means of communication.

Figure 5: Infant deafness rates in Australia



4.5.3 Emerging and micro-community sign languages as endangered

In the developing world, these medical advances have not reached the majority of the population, and there does not appear to be a falling incidence of deafness to quite the same extent. Instead we find a more familiar kind of language endangerment: there are emerging or established macro-community sign languages in some urban areas in these developing countries, but educators setting up Deaf schools have often imposed ASL instead (e.g., Nonaka 2004).

I have personal experience of the situation in Cambodia, for example, where I worked as part of the Deaf Development Programme (see http://www.parish-without-borders.net/ddp/). Here the first Deaf schools were opened in the 1990s. Cambodian Sign Language is an emerging sign language that seems to involve a mixture of Thai Sign Language and various other influences, but it is still a relatively young language because the community has only been forming in the last two decades. Many of the Deaf people I talked to while working in Phnom Penh were first exposed to sign language when they were living in the border camps near Thailand during the Pol Pot era, where they learnt some Thai Sign Language.

There is thus a sign language emerging, but when educators opened the first school for the Deaf, they began trying to teach Deaf children using ASL, dismissing the emerging local sign language. This has happened in other countries in Africa and Asia. In some cases, local sign languages were emerging in the Deaf communities or have been there for some time, but are now being swamped by Western sign languages, particularly ASL. For example, in rural Thailand there is a micro-community village sign language, Ban Khor Sign Language, which has been there for several generations (Nonaka 2004). The large urban centres also have old sign languages: in Bangkok, for example, there was a Deaf community that used Old Bangkok Sign Language, but again educators who established the first schools for the Deaf in Thailand brought in ASL. Modern Thai Sign Language is a mixture of ASL and older Thai signs that existed before the importation of ASL (Woodward 1996).

Thus, many sign languages in developing countries are endangered because another sign language is being adopted as the means of instruction in Deaf education. Because these other languages have not been documented, or may be emerging varieties, educators are often not really aware of them, and have been attracted by the many more resources available for ASL. This approach is well-intentioned (at least, sign language instruction is being adopted!), but it results in local sign language endangerment in many cases.

4.6 Growing awareness of the needs of stakeholders outside the academic community.

The importance of sign language documentation projects for the language sciences is obvious, both for endangered emerging or micro-community sign languages in the developing world, as well as for macro-community sign languages. Despite research having begun on BSL in the late 1970s, much still remains to be discovered about the language. As Brennan (1994) notes:

> It is worth stressing that the amount of research that has been undertaken and is being undertaken with respect to BSL is extremely limited…One of the major difficulties facing sign language linguistics is that many of us are working in 'applied' contexts, before theories and descriptions have been fully elaborated: indeed in some areas, work has not even begun.

Although Brennan made this observation back in 1994, I would argue it is still very much true in 2010. There is a need for more empirical work on BSL phonology, lexicon and grammar, to build on Deuchar (1984), Kyle & Woll (1985) and Sutton-Spence & Woll (1999). There is a need to create a lexical database of the language, so that the lexicon can be better recorded. Only one dictionary organized along linguistic principles currently exists (Brien 1992), and it has fewer than 2000 entries.

However, sign language documentation is not only of interest to the linguistics community, of course – it is also of vital importance as a record of BSL for the British Deaf community. In particular, it is vital to address concerns in British Deaf community about language variation and change (e.g., Elton & Squelch 2008): heritage forms of BSL are not being passed on to a younger generation (especially of young Deaf adults form hearing families) and need to be documented for the future. Currently Deaf researchers and activists are trying to document some of this lexical variation (e.g., Elton & Squelch 2008), but the BSL Corpus Project is the first nation-wide attempt to do so.

5 Sign language documentation projects

The following are some sign language documentation projects that have been recently completed or are currently ongoing:

Completed

1. European Cultural Heritage Online (ECHO) project (BSL, NGT & Swedish Sign Language)[8]: 2003-2004
2. Auslan (Australian Sign Language) Corpus Project[9]: 2004-2007
3. Irish Sign Language (ISL) 'Signs of Ireland' project[10]: 2004
4. Nederlandse Gebarentaal (Sign Language of The Netherlands, or NGT): Corpus NGT[11]: 2006-2008

In progress

1. BSL Corpus Project: 2008-2010
2. Deutsche Gebaerdensprache (German Sign Language, or DGS) Corpus Project: 2009-2023
3. Langue des Signes Malienne (Mali Sign Language, or LSM) 2007-2010
4. Lengua de Señas Mexicana (Mexican Sign Language, or LSM) 2007-2009
5. European Science Foundation EuroBABEL VillageSign project

The Auslan Corpus Project was funded by the Endangered Languages Documentation Programme (ELDP) at SOAS, and was the first major project of its kind in the world. I worked for a short time on the Auslan project with my colleague, Trevor Johnston, along with a number of other hearing and Deaf colleagues. Figure 5 (see above), which shows the falling incidence of deafness, comes from Trevor's work on the Australian Deaf community (Johnston 2004). He was able to argue based on this trend that Auslan is

[8] http://www.let.ru.nl/sign-lang/echo/index.html?http&&&www.let.ru.nl/sign-lang/echo/data.html

[9] http://www.hrelp.org/grants/projects/index.php?projid=48

[10] http://www.tara.tcd.ie/bitstream/2262/1597/1/ITT+paper+vfinal.pdf

[11] http://www.ru.nl/corpusngtuk/

actually endangered. This was perhaps the first time that endangerment was officially recognised for a sign language. The data, which has been deposited in the Endangered Languages Archive (ELAR) at SOAS, will be made accessible to other researchers in the future.

There have been a number of other projects: the Corpus NGT Project, for example, is also groundbreaking and, together with the Auslan Corpus Project, has acted as a model for the BSL Corpus Project (BSLCP). The Corpus NGT Project was the first attempt to create a corpus, to document Sign Language of the Netherlands, and to make an open access archive available online. That project has now reached completion: the data is partly annotated and partly translated into English and Dutch, but it is very much a first phase. Like the Auslan Corpus Project, the aim was to collect the data, make it accessible in an online archive, and then to begin using the data as a resource for future research projects. The BSLCP has a similar aim: we are currently collecting the data, and we hope to create an open access archive. The main difference is that we are actually annotating the data for a number of specific linguistic studies as we collect it.

There are a number of other documentation projects currently being carried out:

- the Mexican Sign Language project led by Claire Ramsey, also funded by ELDP
- the Langue de Signes Malienne (Mali Sign Language) project led by Victoria Nyst, also funded by ELDP
- EuroBABEL VillageSign, an exciting new project that has just been funded by the European Science Foundation, which is going to try to document a number of those micro-community village sign languages, such as Kata Kolok in Bali[12].

[12] There are quite a number of these micro-community sign languages all over the world, and new ones are still being discovered – a Dutch sign language researcher with connections to Suriname recently stumbled upon another village sign language there, for example.

6. The BSL Corpus Project

I am now going to describe in a little more detail the BSL Corpus Project.

6.1 Aims

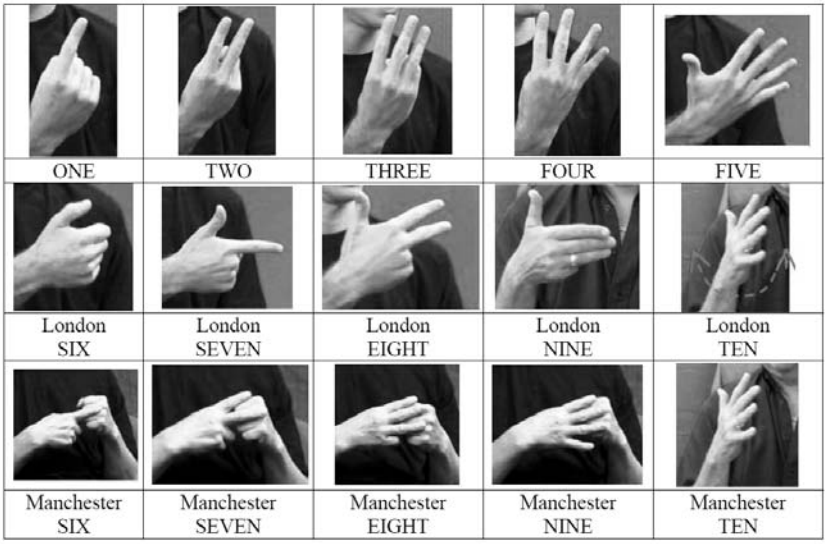
The BSL Corpus Project (BSLCP) aims to create an online, open access corpus of annotated BSL digital video data that will become a shared, peer-reviewable resource and standard reference for BSL researchers and teachers. Full participant consent and metadata (background data about participants etc.) will be included. We are also simultaneously aiming to conduct corpus-based investigations into sociolinguistic variation and change, language contact and lexical frequency.

The reason for the interest in sociolinguistic variation is that it responds to point (d) that Woodbury raised (see above) – namely, linguists becoming aware of the need to work more closely with communities and stakeholders and speakers of these languages because, in the case of endangered languages, the communities are concerned about the future of their languages. The same thing is happening in BSL.

Like many Western sign languages, BSL is a sign language of a macro-community but it has emerged historically out of a collection of micro-community sign languages. In the 19th century, there were about 22 residential schools for the Deaf in different parts of the UK (Kyle & Woll 1985) but, given the lack of social mobility at the time, Deaf people did not travel around very much or meet Deaf people from other regions. The UK is a comparatively small country, but there was very little mobility, so the result was quite distinct varieties of BSL emerged in each of these local Deaf micro-communities.

It has perhaps only been since the late 20th century that we have seen an emergence of a national Deaf identity and a macro-community in the UK (many signers in Bristol, for example, reported difficulty understanding signers from other parts of the country as recently as 1980, see Kyle & Allsop 1982). A lot of traditional regional lexical differences across the country still exist, but some variants are only used by the older generation of British signers. The British Deaf community is very concerned about documenting and preserving the rich variation before it possibly disappears – there appears to be dialect levelling and standardisation within BSL because of increased mobility and the emergence of a national British Deaf identity.

Figure 6: Dialect variation in BSL



In Figure 6, there is an example of some of the regional variation: there are several different number systems in BSL, as work currently undertaken by the BSL Corpus Project led by Rose Stamp is showing. In London and Manchester, the traditional number signs for ONE to FIVE are identical (and the same as hearing people's number gestures), as shown below, but the numbers SIX to TEN are quite different (note that London TEN is distinguished from FIVE by an additional repeated twisting movement, represented by the red arrow here). The Manchester number sign system is distinct, unlike any used elsewhere in the UK. It is one of the few base-5 systems documented in any sign language: a closed fist represents 5 and numbers above 5 are produced by using the same signs as 1 to 4 and placing them on the fist; two fists together represents 10.

6.2 Methodology: Sociolinguistic approach

BSLCP data collection is being carried out in eight of the largest urban centres across the UK: London, Bristol, Birmingham, Manchester, Newcastle, Cardiff, Glasgow and Belfast. Unlike many of the other corpus projects (but like sociolinguistic variation projects conducted on ASL, Auslan and NZSL – see Lucas et al. 2001; Schembri et al. 2009), we are using an approach that is very much driven by sociolinguistic methodology. We plan to film at least 30 Deaf native and near-native signers of BSL (i.e. Deaf people who have been

exposed to BSL by 7 years of age and/or from birth) who are long-term residents in each of the eight sites (they must have lived in the same region for the last ten years of more). Individuals are only filmed with others from the same region. We balance the selection of participants for:

- **age**: we attempt to recruit equal numbers of participants in four age groups: (a) 18-35 (b) 36-50, (c) 51-65 and (d) over 65, and we try to film all participants in a pair in which they are matched with someone in the same age group.
- **gender**: roughly equal numbers of men and women in all age groups.
- **ethnicity**: although we aim only to film British-born Deaf people, we will include 10% of participants from a South Asian or Afro-Caribbean ethnic background, reflecting the UK population distribution in the 2001 Census.
- **language background**: we are trying to ensure that about a third of participants are Deaf people with Deaf parents (who have therefore grown up with sign language from birth), and two thirds are Deaf people from hearing families (who have been exposed to sign language and learned it from peers at school).
- **social class**: participants are drawn from a range of educational and occupation backgrounds, although socioeconomic stratification is less marked in the British Deaf community due to the relative recent emergence of a professional class.

6.3 Methodology: Recruitment and data collection

Another aspect of our approach with the BSLCP, very much influenced by sociolinguistics work on accommodation and style shifting (Lucas & Valli 1992), is that filming occurs without any hearing people present (including me, the director of the project, because I am a hearing person). Lucas & Valli (1992) showed that at least some Deaf American signers accommodate their language production to the needs of hearing signers. In particular, they shift towards more English-like varieties of sign language, use more mouthing of English words, and follow English word order when hearing people are present. In order to avoid this, and to try to document the vernacular variety of BSL as much as possible, we do not have any hearing people present during filming. Furthermore, like other sociolinguistic variation projects (e.g., Lucas et al. 2001), recruitment of the 30 people in each site is carried out by a paid

Deaf community fieldworker[13]. The fieldworker has local Deaf community knowledge that is invaluable for approaching and recruiting potential participants, and for ensuring that the participants that are filmed together are family members, friends or acquaintances that feel comfortable with each other.

We go to the recording site over two or three different visits and film there. We have specific requirements for filming to ensure that we can adequately see and annotate the sign language production data captured on the videotapes:

- we use two lights: this is to ensure good quality lighting in the room which can vary from one filming location to the next (we film mostly in Deaf clubs and associations, of which there may be two or three at each site)
- we film people in pairs: larger groups require more cameras, or would mean that the images of each individual signer would be smaller on the screen
- we use three to four high-definition digital video cameras (one focused on participant A, one focused on participant B, one focused on the pair, and sometimes the fourth camera focuses on the field worker).

We collect data of four types:

- personal anecdotes from each individual.
- 30 minutes of free conversation on any topic(s) chosen by the interviewees.
- interviews where we ask questions about BSL usage, variation, and attitudes towards the language (these interviews are conducted by the Deaf fieldworker).
- elicitation of signs for 100 key concepts which are expressed in BSL using lexical signs that are known to vary across the country.

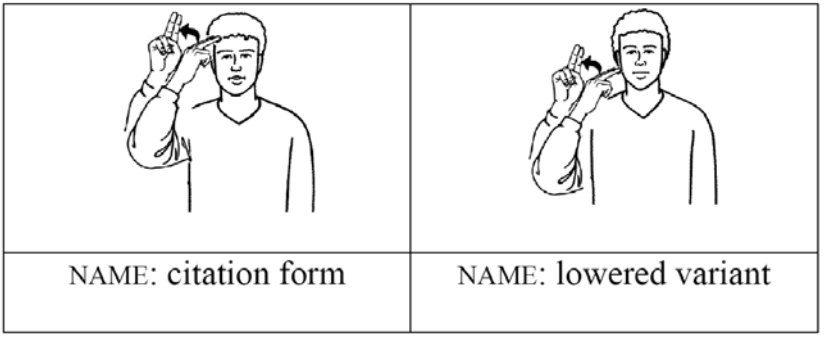
We also plan to collect additional data, although the details of this second phase of data collection have not yet been finalised at the time of writing.

---

[13] We have eight different fieldworkers working in eight different sites who recruit 30 participants matching our criteria; they are then assembled at a specific point for filming.

We use two blue background screens in order to maximise our ability to code the subtleties of the sign language data – pale colours in the background, for example, make it very difficult to see the handshapes of fair-skinned signers. Two lights are used, as mentioned above, one placed near each participant, but out of shot of the video camera.

Figure 7: Sign variants of NAME



We seat the participants in chairs without arms to prevent them from resting their elbows while signing as this interferes with sign language production, especially the production of signs on their usual locations – people tend not to hit target locations that are higher on the body when they are resting their elbows. For example, the BSL sign NAME is normally produced at the forehead (see Figure 7), but in casual signing it can be produced at lower locations.

If an interviewee is leaning on the arms of a chair, then they are possibly more likely to use the casual variant. We want to know about variation in signs due to formality and other sociolinguistic factors, and do not want physical features of the interview layout to interfere with natural patterns.

Figure 8: Clothing restrictions

Also, we insist that all participants wear plain coloured clothing, as shown in Figure 8 (two signers from Glasgow). We bring along back-up plain dark t-shirts in case interviewees turn up wearing something inappropriate. Again, coloured clothing can interfere with the ability to code subtle features of signs, such as finger configurations. This requirement is based on unfortunate experiences in a previous Auslan project when we filmed an elderly couple in Queensland, Australia. They were really excited about being filmed together and so they wore their favourite shirts to the filming session: Hawaiian shirts with a variety of purple, red and orange hibiscus flowers all over them. Like other elderly white Australians from Queensland, their skin was freckly and pink (from a lifetime's exposure to sub-tropical sun), and wearing colourful Hawaiian shirts with this skin colour is not at all a good combination for coding purposes. So, much to their disappointment, we asked them to replace their shirts with plain shirts.

6.4 Issues with open access archives

There are a number of ethical and methodological issues associated with the creation of open access archives for sign language documentation. Crasborn (2008) discusses several of them, such as the protection of participants' privacy. Privacy protection is a very difficult matter, given that signers can be identified by those who know them in the video data, and will be even more difficult in the future when facial recognition technology becomes more widely available. There are also possible normative effects of documentation on usage within the Deaf community. It may be that signers will come to view the collection not as a record of actual sign language use, but as some kind of set of norms for BSL usage, merely as a result of its existence as an online resource that can be consulted. What I would like to focus on here, however, relates to a key issue for sign language researchers: what kind of data is collected for an open-access corpus?

6.4.1 The vernacular variety

As Tagliamonte (2006) explains, a specific aim of sociolinguistics is to study the 'vernacular' variety. This is the variety used by speakers when they are monitoring their language use least closely (Labov 1972), or the spontaneous speech reserved for intimate or casual situations which is taken to reflect the most systematic form of the language acquired by the speaker, prior to any subsequent efforts at hypercorrection or style shifting (Poplack 1993).

In Figure 9, I present a cline of language varieties used in the British Deaf community. Each of these varieties represents an abstraction, but these categories are similar to widely used categories in sign language literature

generally (Sutton-Spence & Woll 1999; Lucas & Valli 1992). Unlike other work, however, I have added 'BSL self-conscious style' at the left end of the continuum for varieties in which Deaf bilinguals consciously reduce their English code-mixing to the best of their ability, either to achieve particular linguistic effects, as in sign language poetry, or to reflect attitudes related to linguistic purism and prescriptivism (cf. Wertheim 2006 on Tartar and Russian code-mixing). Observation suggests that 'vernacular BSL' and 'contact signing' are perhaps the varieties most commonly used in most informal situations by Deaf bilinguals, although mixed sign language in which English is clearly the matrix language is also not uncommon. The aim of the BSL Corpus Project is to collect as much data in vernacular BSL varieties as possible; this is the reason for the methodology described above (no hearing people present, filming only native and early learners of BSL, pairing up individuals who are comfortable being filmed together, and so on).

Figure 9: Continuum of styles

Style	**BSL self-conscious style**	**Vernacular BSL**	**Contact signing**	**Signed English/ Sign Supported English**	**Spoken English**
Characteristics	Minimal English influence	English code-mixing, but BSL is the matrix language	Code-mixing between English and BSL, with neither as matrix	English as the matrix language	Some, little or no BSL influence

6.4.2 The Observer's Paradox

As Labov (1972) argued, the main aim of linguistic research must be to find out how people talk when they are not being systematically observed, yet we can only obtain this data by systematic observation. This is the Observer's Paradox.

As a sociolinguist, I am really interested in the variety of BSL that people use when they are with Deaf friends and family with whom they feel comfortable and relaxed. I assume that this vernacular variety is the most systematic, the least influenced by hypercorrection and self-conscious monitoring, and the variety which the signer acquired from birth (if they have Deaf parents), or from a very young age (if they learned the language from Deaf peers at school). Despite all the cameras, lights, blue background screens, clothing requirements, we try to ensure that our participants are as relaxed as possible.

We also try to reassure them about the nature of the project. We use a very detailed video consent form, where we outline everything we are going to do with the data, including putting it on the Internet as part of an open access archive. We make sure that people consent to all of those uses before they come to the filming. A key job for the fieldworkers when they are recruiting is to make it really clear to the Deaf participants what they will be agreeing to by participating[14].

However, we place limits on how well participants should know their conversational partner. We do not film married couples, especially elderly, retired married couples, because we have found that it is difficult for them to sustain 30 minutes of conversation. Since they typically see each other every day, they do not have a great deal of new things to say to each other. (In the previous Australian project one elderly couple ran out of things to say to each other after ten minutes of the 30-minute data collection session. They got up, walked away from the cameras and said to the field worker "we've run out of things to say, what do we do next?" The fieldworker told them to go back, sit down, and think of something to talk about. They did so and the husband asked his wife "what's your name?", "how many children do you have?", "what are their names?", "how old are they?", "do you have grandchildren?" None of this is naturalistic conversation!) So, when filming couples, we try to split them up, for example, by pairing off a husband and wife with other people separately.

Although we do as much as we can to create a relaxed environment for participants and to encourage the use of vernacular varieties of BSL, it is clear that there is also a self-selection process at work. People who agree to participate may be those who are already comfortable with being filmed. It is possible that we are going to have a collection of signing from 240 Deaf people who use the language in a way that is not representative because, even though there is a primary addressee right there in their conversation partner, participants might also be thinking "I don't know who else is going to see this, so should I adjust my signing accordingly?".

---

[14] We have had many potential interviewees decide not participate for this reason. Many were not comfortable with the open-access nature of the collection we wish to create. In particular, elderly people who do not use the Internet did not know what it is and were really unsure about consenting to putting data on the web.

6.4.3 Bell's audience design and participant roles

The anthropologist and sociolinguist Suzanne Wertheim (2006) has proposed that we can use the work of Allan Bell to better understand the observer's paradox and the challenge this poses for language documentation and description (Wertheim was working on a spoken language, Tartar, but I will show how her proposal works equally well for sign languages). Bell (1984: 197) claims that "…at all levels of language variability, people are responding to other people. Speakers are designing their style for their audience". He identifies five 'participant roles' for any speech situation (ordered from most to least effect on the variety of language produced):

- **Speaker** (1st person): the person who is speaking
- **Addressee** (2nd person): the person or people addressed by the speaker
- **Auditor** (3rd person): people not addressed by the speaker, but referred to
- **Overhearer** (known to be listening): unratified to participate in the speech event, and neither addressed nor referred to.
- **Eavesdropper** (unknown): unratified, unknown and not referred to.

These participant roles can be adapted for sign language events, but I suggest adding one more participant role: the '**viewer** of video data' (see Figure 10). Wertheim (2006: 723) suggests that in spoken language documentation, the "unusual audience attributes of the end listener of recorded speech, highlighted by the act of recording and the presence of recording equipment, can cause this audience role to be of primary salience and focus, regardless of the fieldworker's actual participant role in the speech event". The same must also be true of the recorded sign language events in the BSL Corpus Project collection. Participants do not know who the end users of the collection will be: they may conceive of them as the BSL Corpus Project team in combination with other known sign language researchers, or as researchers and members of the Deaf community, or as Deaf and hearing people completely unknown to any of the participants. How they may or may not design their sign production for this unknown audience is also unknown.

Figure 10: Types of participants

	Known	*Ratified*	*Addressed*	*Person*
Addressee	+	+	+	2nd
Viewer	+	+	-	3rd
Unratified viewer	+	-	-	N/A
Eavesdropper	-	-	-	N/A
Viewer of video data	-	+	-/+	?

6.4.4 Critiques of sociolinguistic assumptions

We need to be aware of the audience design issue discussed by Wertheim (2006), but she also suggests that 'performed' or 'self-conscious' speech produced for fieldworkers is important for systematic linguistic analysis and in gaining insights into local language ideologies and linguistic norms. Indeed, the sociolinguist Schilling-Estes (1998) found that 'performance speech' in the Ocracoke dialect of English is regularly patterned, unlike what has been claimed by Labov (1972), Tagliamonte (2006), and Poplack (1993). Thus, although signers may vary in the degree of self-consciousness and language mixing they exhibit in their sign language production recorded for the BSL Corpus Project, this does not necessarily undermine the usefulness of the resulting data.

6.4.5 Dual access BSL Corpus

Despite the reassuring findings from Wertheim (2006) and Schilling-Estes (1998), one of the ways that we have tried to overcome concerns about the observer's paradox is to exclude the 30-minute conversations from the planned open access archive. This subsection of the collection will have restricted access, perhaps only available to researchers, who will have to sign a confidentiality agreement. That is, there will be a dual access archive, with the interview data, lexical elicitation tasks and narratives all made open **access**, but the conversational data given restricted access (see Nathan's discussion of access in the archiving chapter of this volume).

We have explained this arrangement to Deaf participants to encourage them to feel more comfortable. One of the reasons we have opted to do this with the free conversation data is because we know what people often do when they get together and have conversations – they talk about themselves and other people they know. In fact, a lexical frequency study conducted on 100,000 signs of a corpus of NZSL (McKee & Kennedy 2006) found that the most frequent lexical item in the corpus was the sign PRO-1 (e.g. meaning 'I' or 'me') and the second most common was PRO-3. Deaf people, like everyone else, like to talk about themselves and other people. We do not wish to prevent this from happening, because some of the most vernacular varieties of sign language might be used when participants are relaxed, gossiping about other people and thus engrossed in their conversation. We know, however, that making this kind of data widely available on-line may be a sensitive issue. We have attempted to make it clear to our participants that this free conversation data will not be open-access and thus will not be available for all to see.

6.4.6 Other challenges

There is not space here to discuss many of the other challenges, which confront sign language documentation work in general, and the BSL Corpus Project in particular. One of the most important issues has been raised in Johnston (2010), which points out that, although ELAN and related software has made large-scale sign language documentation projects possible, it does not solve a very basic issue in sign language representation. One requirement of a sign language corpus is that it should be machine-readable (i.e. able to be manipulated by computers), but only a systematic approach to annotation that involves lemmatisation of the sign language glosses can make sign language corpora machine-readable. In corpus linguistics, lemmatisation is the process of grouping together different inflected forms of a word (e.g., *walks*, *walked, walking*) so they can be analysed as members of a single item (e.g., *walk*). In sign language documentation, the translation of the sign language lexical items has to be lemmatised as well, so that all variants of a related sign can be identified and analysed as a single sign. In the case of the Auslan corpus that Johnston (2010) discusses, this lemmatisation process is made more straightforward by the existence of a comprehensive lexical database for Auslan (Johnston 2001). No such BSL lexical database currently exists, so our lemmatisation work will have to be done on the fly, at least until funding for a BSL lexical database can be obtained (currently, we have proposed this in a new project to begin in 2011).

7. Conclusion

To summarise, I began this discussion about the relatively new field of documentary sign linguistics with some background about sign languages, before moving on to describe how a number of key changes have seen the emergence of documentary sign language linguistics. These included developments in technology of data representation such as ELAN; a new attention to linguistic diversity both in and outside of sign linguistics; concern about sign language and Deaf community endangerment; and a desire to address some of the concerns of the Deaf community about documenting sign language variation and change. I then discussed examples of sign language documentation projects, with a specific focus on the BSL Corpus Project. In particular, I explored some of the key issues in sign language documentation methodology, such as the implications of the observer's paradox for data collection and the creation of open-access archives of video data.

References

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2), 145-204.

Brennan, Mary. 1994. Word order: Introducing the issues. In Mary Brennan & Graham H. Turner (eds.), *Word-order issues in sign language: Working papers presented at a workshop held in Durham, 18-22 September 1991*, 9-45. Durham: International Sign Linguistics Association.

Brien, David (ed.). 1992. *Dictionary of British Sign Language/English*. London: Faber & Faber.

Collins-Ahlgren, Marianne. 1989. *Aspects of New Zealand Sign Language.* Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.

Corbett, Greville G. 2006. *Agreement* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Crasborn, Onno. 2008. Open access to sign language corpora. In Onno Crasborn, Eleni Efthimiou, Thomas Hanke, Ernst D. Thoutenhoofd & Inge Zwitserlood (eds.), 3rd workshop on the representation and processing of sign languages: Construction and exploitation of sign language corpora [a workshop given at the Sixth International Conference on Language Resources and Evaluation, 26 May – 1 June 2008, Marrakech, Morocco], 33-38. Paris: European Language Resources Association.http://www.lrec-conf.org/proceedings/lrec2008/ (28 January 2010).

de Beuzeville, Louise, Trevor Johnston & Adam Schembri. 2009. The use of space with indicating verbs in Auslan: A corpus-based investigation. *Sign Language & Linguistics* 12(1): 53-82.

Deuchar, Margaret. 1984. *British Sign Language* (Language, Education and Society). London: Routledge & Kegan Paul.

Elton, Frances & Linda Squelch. 2008. *London and South East regional signs.* London: Lexisigns.

Emmorey, Karen. 2002. *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5), 429-448.

Fischer, Susan D. 1978. Sign language and creoles. In Patricia Siple (ed.), *Understanding language through sign language research* (Perspectives in Neurolinguistics and Psycholinguistics), 309-331. New York: Academic Press.

Folven, Raymond J., John D. Bonvillian & , Michael D. Orlansky. 1984. Communicative gestures and early sign language acquisition. *First Language* 5(14), 129-143.

Fromkin, Victoria, Robert Rodman & Nina Hyams. 2006 *An introduction to language*, 8th edn. Boston, MA: Thomson Wadsworth.

Guerra Currie, Anne-Marie P., Richard P. Meier & Keith Walters. 2002. A crosslinguistic examination of the lexicons of four signed languages. In Richard P. Meier, Kearsy Cormier & David Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, 224-236. Cambridge: Cambridge University Press.

Hoffmeister, Robert. 2000. A piece of the puzzle: ASL and reading comprehension in Deaf children. In Charlene Chamberlain, Jill P. Morford & Rachel I. Mayberry (eds.), *Language acquisition by eye*, 143-163. Mahwah, NJ: Lawrence Erlbaum Associates.

Jeanes, Raymond C. & Brian E. Reynolds. 1982. *Dictionary of Australasian signs for communication with the deaf*. Melbourne: Victorian School for Deaf Children.

Johnston, Trevor. 2001. The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics* 4(1-2), 145-169.

Johnston, Trevor. 2003. BSL, Auslan and NZSL: Three signed languages or one? In A. Baker, B. van den Bogaerde & O. Crasborn (eds.), *Cross-linguistic perspectives in sign language research: Selected papers from TISLR 2000* (International Studies on Sign Language and Communication of the Deaf 41), 47-69. Hamburg: Signum Verlag.

Johnston, Trevor. 2004. W(h)ither the deaf community? Population, genetics, and the future of Australian Sign Language. *American Annals of the Deaf* 148(5), 358-375.

Johnston, Trevor. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15(1), 104-129.

Johnston, Trevor & Adam Schembri. 2007. *Australian Sign Language (Auslan): An introduction to sign language linguistics.* Cambridge: Cambridge University Press.

Kegl, Judy, Anne Senghas & Marie Coppola. 1999. Creation through contact: Sign language emergence and sign language change in Nicaragua. In Michael DeGraff (ed.), *Language creation and language change:*

Creolization, diachrony, and development (Learning, Development, and Conceptual Change), 179-237. Cambridge, MA: MIT Press.

Kendon, Adam. 1988. *Sign languages of aboriginal Australia: Cultural, semiotic and communicative perspectives*. Cambridge: Cambridge University Press.

Klima, Edward & Ursula Bellugi. 1979. *The signs of language*. Cambridge, MA: Harvard University Press.

Kyle, J. G. & B. Woll. 1985. *Sign language: The study of deaf people and their language*. Cambridge: Cambridge University Press.

Kyle, J. G. & L. Allsop. 1982. *Deaf people and the community: Final report to the Nuffield Foundation*. Bristol: University of Bristol, School for Education.

Labov, William. 1972. *Language in the inner city: Studies in the Black English vernacular* (Conduct and Communication). Philadelphia, PA: University of Pennsylvania Press.

Lane, Harlan. 1984. *When the mind hears: A history of the Deaf*. New York: Random House.

Liddell, Scott K. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Lucas, Ceil & Clayton Valli. 1992. *Language contact in the American deaf community*. San Diego, CA: Academic Press.

Lucas, Ceil, Robert Bayley & Clayton Valli. 2001. *Sociolinguistic variation in American Sign Language* (Sociolinguistics in Deaf Communities). Washington, DC: Gallaudet University Press.

Marschark, Marc & Patricia Elizabeth Spencer. What we know, what we don't know, and what we should know. In Marc Marschark & Patricia Elizabeth Spencer (eds.). 2003. *Oxford handbook of deaf studies, language, and education*, 491-494. Oxford: Oxford University Press.

McKee, David & Graeme Kennedy. 2006. The distribution of signs in New Zealand Sign Language. *Sign Language Studies* 6(4), 372-390.

Meir, Irit, Carol A. Padden, Mark Aronoff & Wendy Sandler. 2007. Body as subject. *Journal of Linguistics* 43(3), 531-563.

Mitchell, Ross E. & Michael A. Karchmer. 2004. Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies* 4(2), 138-163.

Nonaka, Angela M. 2004. The forgotten endangered languages: Lessons on the importance of remembering from Thailand's Ban Khor Sign Language. *Language in Society* 33(5), 737-767.

Padden, Carol A. 1988. *Interaction of morphology and syntax in American Sign Language* (Outstanding Dissertations in Linguistics). New York: Garland.

Petitto, Laura Ann. 2000. On the biological foundations of human language. In Karen Emmorey & Harlan Lane (eds.), *The signs of language revisited: An anthology in honor of Ursula Bellugi and Edward Klima*, 449-473. Mahwah, NJ: Lawrence Erlbaum Associates.

Pinker, Steven. 1994. *The language instinct*. New York: W. Morrow and Co.

Poplack, Shana. 1993. Variation theory and language contact. In Dennis R. Preston (ed.), *American dialect research: Celebrating the 100th anniversary of the American Dialect Society, 1889-1989* (Centennial Series of the American Dialect Society), 251-286. Amsterdam: Benjamins.

Rubino, Francesco, Allan B. Hayhurst, Josif Guejlman & Willard J. Madsen (World Federation of the Deaf, Unification of Signs Commission). 1975. *Gestuno: International sign language of the deaf: The revised and enlarged book of signs agreed and adopted by the Unification of Signs Commission of the World Federation of the Deaf*. Carlisle: British Deaf Association.

Sandler, Wendy & Diane Lillo-Martin. 2006. *Sign language and linguistic universals*. Cambridge: Cambridge University Press.

Schembri, Adam, David McKee, Rachel McKee, Sara Pivac, Trevor Johnston & Della Goswell. 2009. Phonological variation and change in Australian and New Zealand Sign Languages: The location variable. *Language Variation and Change* 21(2), 193-231.

Schilling-Estes, Natalie 1998. Investigating 'self-conscious' speech: The performance register in Ocracoke English. *Language in Society* 27(1), 53-83.

Stokoe, William C. 1960. *Sign language structure: An outline of the visual communication systems of the American Deaf* (Studies in Linguistics, Occasional Papers 8). Buffalo, NY: University of Buffalo.

Sutton-Spence, Rachel & Bencie Woll. 1999. *The linguistics of British Sign Language: An introduction*. Cambridge: Cambridge University Press.

Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation* (Key Topics in Sociolinguistics). Cambridge: Cambridge University Press.

Wertheim, Suzanne. 2006. Cleaning up for company: Using participant roles to understand fieldworker effect. *Language in Society* 35(5), 707-727.

Woll, Bencie & Paddy Ladd. 2003. Deaf communities. In Marc Marschark & Patricia Elizabeth Spencer (eds.), *Oxford handbook of deaf studies, language, and education*, 151-163. Oxford: Oxford University Press.

Woodbury, Tony. 2003. Defining documentary linguistics. In Peter K. Austin (ed.), *Language documentation and description*, vol. 1, 35-51. London: SOAS.

Woodward, James. 1996. Modern Standard Thai Sign Language, influence from ASL, and its relationship to original Thai Sign Language varieties. *Sign Language Studies* 92: 227-252.

Discussion Questions

1. Exactly how is ELAN (and related annotation software, such as ilex) used as part of sign language documentation projects? What standards for annotation currently exist? And what standard metadata categories have been proposed for sign language documentation projects?

2. This chapter discussed the debate in the sign language linguistics literature about the nature of indicating verb signs, and mentioned how sign language corpus projects have begun to contribute new kinds of evidence to the debate. What were the key findings of the Auslan corpus study conducted by de Beuzeville, Johnston and Schembri (2009)? What other contributions to an understanding of sign language structure and use might sign language documentation projects make in the future?

3. Apart from the Observer's Paradox and the nature of open-access archives, what additional ethical and sociolinguistic considerations are important in sign language documentation projects? How do these compare to ethical issues in spoken language documentation?