# Applying lexicostatistical methods to sign languages: How *not* to delineate sign language varieties.[1]

Nick Palfreyman
(nickpalfreyman@cantab.net)

International Institute for Sign Languages and Deaf Studies,
University of Central Lancashire.

The reliability and general efficacy of lexicostatistical methods have been called into question by many spoken language linguists, some of whom are vociferous in expressing their concerns. Dixon (1997) and Campbell (2004) go as far as to reject the validity of lexicostatistics entirely, and Dixon cites several publications in support of his argument that lexicostatistics has been 'decisively discredited', including Hoijer (1956), Arndt (1959), Bergslund and Vogt (1962), Teeter (1963), Campbell (1977) and Embleton (1992). In the field of sign language research, however, some linguists continue to produce lexicostatistical studies that delineate sign language varieties along the language-dialect continuum. Papers referring to lexicostatistical methods continue to be submitted to conferences on sign language linguistics, suggesting that these methods are – in some corners of the field, at least – as popular as ever.

Given the ongoing popularity of lexicostatistical methods with some sign language linguists, it is necessary to be as clear as possible about how these methods may generate misleading results when applied to sign language varieties, which is the central aim of this paper. Furthermore, in several cases, lexicostatistical methods seem to have been deployed for a quite different purpose – that of establishing mutual intelligibility – but the suitability of lexicostatistical methods for such a purpose has not been openly addressed, and this is dealt with in Section 4.

This paper is based on a review of the literature and was written following my time living in Indonesia (2007-2009) and working on research with users of Indonesian sign language varieties (between 2010 and 2013). Section 1 outlines the history of lexicostatistical methods in linguistics, while Section 2 is concerned with lexicostatistical studies of sign language varieties. Problems in applying lexicostatistical methods to sign languages are set out in Section 3, and in Section 4 I discuss the stated and implicit aims of sign language sociolinguists in making recourse to lexicostatistical methods. The final section highlights the need for sign language sociolinguistics to move away from lexicostatistics as a false proxy for mutual intelligibility.

## 1.  What is lexicostatistics?

Lexicostatistics is a method of classification that entails comparing the vocabulary of different language varieties to find a measure of distance through the application of a statistical scale, and enables linguists to reconstruct family-trees for groups of languages that are known to be related (Embleton, 2000; Campbell, 2004). Perhaps the most well-known proponent of lexicostatistics is Morris Swadesh (1950,

---

[1] This paper is based on pp. 17-20 and pp. 22-55 of my revised PhD thesis, *Sign language varieties of Indonesia: Linguistic and sociolinguistic perspectives* (University of Central Lancashire) which was accepted with minor amendments in 2014. I would like to thank Ulrike Zeshan, Connie de Vos, Sheila Embleton and David Gil for discussing lexicostatistics with me, and my PhD examiners for their feedback. Any errors remain my own.

1954, 1955) – although several linguists proposed and developed lexicostatistical methods prior to this (Embleton, 2000). The application of Swadesh's methods have been controversial from the outset: for an early example, see criticisms made by Gudschinsky (1956b) concerning Lees (1953).

Several linguists have responded to the perceived methodological weaknesses of lexicostatistics individually through a series of modifications (see Trask, 1996, and Embleton, 2000, for examples); an alternative 'pick and mix' approach has seen linguists adopt some elements of lexicostatistics and jettison others, in accordance with their own objectives. Given the plethora of individual methodological combinations used in the literature, there is no single uniform, commonly recognised 'lexicostatistical method', and some working definitions are essential here to avoid confusion.

The first useful distinction is between *classical lexicostatistics* and *preliminary lexicostatistics*. This distinction is proposed by Starostin (2010), who stipulates that classical lexicostatistics is only conducted once a historic relationship between the language varieties in question has already been demonstrated. Classical lexicostatistics is the last stage in a long process of determining the historical relationships between a series of language varieties. Preliminary lexicostatistics, on the other hand, entails the use of lexicostatistical methods before any relationship between the languages has been determined. This approach is open to accusations of circular reasoning: lexicostatistical methods are used to establish that there *is* a relationship, and on that basis, the same methods determine the nature of that relationship.

Central to both methods is the concept of the 'cognate', which can best be explained with recourse to an example. Historical linguists consider German and English to be related because they have descended from a single language variety – West Germanic – in use around 2000 years ago (Hawkins, 2009). As a result of this common origin, the German word *Tanz* and the English word *dance* came from the same proto-form, and so they are cognate. As the West Germanic variety split – see Figure 1 – the proto-form changed in different ways as, through time, its sounds became subject to different but regular phonological processes in each 'descendent' language variety. Regular sound correspondence between German [t] and English [d] means that *Tanz/dance* is not the only cognate pair: others include *Tag/day, tot/death*, *Tür/door* – and *gut/good*, where the affected consonant is in a final position (Ratcliffe, 1998:14).
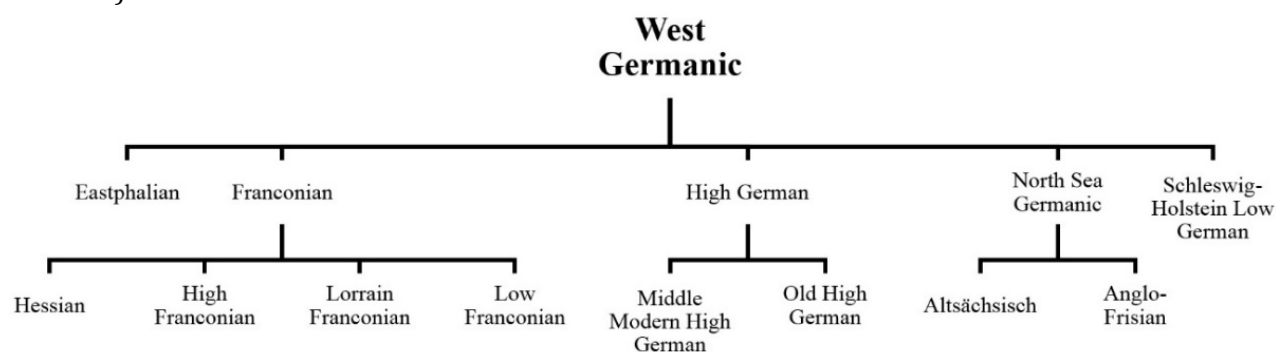


**Figure 1.** A section of the West Germanic family tree (based on Hammarström et al., 2014).

Written documentation is especially useful in permitting historical linguists to use techniques such as regular sound correspondence in order to identify cognates; methods from classical lexicostatistics can then be applied. One of the problems is that it can be difficult if not impossible to identify cognates simply by finding formal similarities. To give but one example, Spanish *mucho* and English *much* are formally and semantically similar, but not cognate: *mucho* comes from the Latin *multum* meaning 'much', while *much* comes from Old English *micel* meaning 'big' (Warnow, 1997:6586). Conversely, French *chef*

and English *head* are cognate, even though this is no longer formally apparent (Embleton, 2000:149).[2] It is primarily for this reason that preliminary lexicostatistics has received criticism: identifying cognates on the basis of formal similarity alone is a risky endeavour.

Another reason for caution is that formal similarities can arise through contact between language varieties in the period following a language 'split'. There are formal similarities in pairs such as the French *famille* and the English *family*, for example, but these are *not* treated as cognate; rather, they occur as a result of language contact. In that case, the Norman invasion of England in 1066 led to many borrowings from language varieties across the English Channel. Embleton (2000:149) points out that the 'splits' manifest in the *Stammbaum* or 'family-tree' model of language change, where language varieties separate and change in different ways, are often inaccurate. Varieties do not always separate quickly and 'cleanly', nor do they always subsequently develop independently of one another. Conversely, it is very common for proximate language varieties to borrow from each other.

Having established a historical relationship between language varieties, the classical lexicostatistics method requires linguists to ascertain how many items on a fixed word list are 'cognates' for these varieties. Once a percentage is obtained, this is applied to a classificatory scale in order to group these varieties. The resulting percentage of cognates – between English and German, for example – depends on the items that are examined. Importantly, lexicostatisticians have argued that the core or basic vocabulary must be used, because it is (apparently) more resistant to phenomena such as borrowing than are the peripheral or general vocabulary (Gudschinsky, 1956a:613; Crowley, 1997:171). Lexicostatistics also rests on the assumption that the rate of lexical replacement is more or less stable (Crowley, 1997:172), and Swadesh (1950) developed a list of 'core vocabulary' comprising noncultural lexical items that are supposedly less prone to borrowing.[3]

All of these assumptions have subsequently been questioned by linguists. The very notion that there is such a thing as a basic or core vocabulary of items, that are independent of language or culture, is criticised by Campbell (2004:201), who also doubts the validity of the assumption that the rate of lexical retention can be constant through time, and that the rate of loss is the same cross-linguistically (202). There are several documented examples where basic vocabulary changes rapidly and unevenly, through borrowing and other phenomena, and this distorts the results, leading Embleton (2000) to describe the rate of lexical retention assumption as a 'grossly simplifying' one.

Several different scales are proposed in the literature for the classification of language varieties, which can lead to confusion (Crowley, 1997:184), though most seem to be a variation on the one proposed by Swadesh (Gudschinsky, 1956a:621) shown below in Table 1.[4] It seems that the linking of thresholds (81%, 36% and so on) with class names ('language', 'family', 'stock') was arbitrary, which is particularly striking given how widely this scale has been used. Furthermore, Crowley (1997:173) carefully notes that the term 'family' is used here with a different meaning: 'lexicostatisticians are using the term *family* in a completely different way from the way [it is commonly used].' Historical linguists usually take 'language family' to mean all language varieties that have descended from a common parent language, regardless of how close or distant the relationships with each other. According to Crowley (1997:173),

---

[2] Parallels can be found for signed languages: there are formally similar (iconic) signs in Japanese Sign Language and Mexican Sign Language, which are unrelated languages (Currie, Meier & Walters, 2002); conversely, Frishberg (1975) gives examples of signs in ASL that derive from eighteenth-century French Sign Language; diachronic changes have altered the forms of some of these signs in ways that may prevent them from being identified on formal grounds as cognate with contemporary LSF signs.

[3] Initially this list comprised 200 items, but was later reduced to 100 (Swadesh, 1971:283).

[4] Although Gudschinsky (1956a:621) cites Swadesh (1954), she unfortunately omits this reference from her bibliography. The article that she refers to appears to have been published by Swadesh in *Word*, but this journal is no longer in publication and I have been unable to find back issues. It would be interesting to see the reasoning that Swadesh gives for his scale, not least given the impact it has made on sign language sociolinguistics.

lexicostatisticians use the term *family* simply as 'a particular level of subgrouping in which the members of that subgroup share more than 36% of their core vocabulary'. This is confusing, and alternative labels for sub-groupings would surely be more appropriate.[5]

*Table 1. Terms for categorising languages suggested by Swadesh (cited by Gudschinsky, 1956a:621).*

| term | divergence (centuries) | cognate (percent) |
|---|---|---|
| language | 0-5 | 100-81 |
| family | 5-25 | 81-36 |
| stock | 25-50 | 36-12 |
| microphylum | 50-75 | 12-4 |
| mesophylum | 75-100 | 4-1 |
| macrophylum | over 100 | less than 1 |

A related method, glottochronology, aims to assign a date to the separation of languages, and builds upon the assumptions of lexicostatistical methods. Trask (1996:362) and Embleton (2000:145f) note that 'lexicostatistics' and 'glottochronology' are used interchangeably by some linguists, and I restrict the use of the term glottochronology here to refer to the techniques that seek to analyse time depth, whereby 'the greater the time depth which separates the members of a language family from their common ancestor the greater the degree of differentiation between them' (Bynon, 1977:267). Glottochronology is effectively an extension of classical lexicostatistics, but is more controversial, and doubts as to its premises – strong doubts, in some cases – are expressed by Trask (1996), Embleton (2000), Campbell (2004) and Joseph and Janda (2007), among others.

Briefly, glottochronology seeks to pinpoint the date when a proto-language split by comparing the reconstructed proto-language with its 'descendant' language. Lees (1953) attempts to quantify a standard rate of change in terms of morpheme decay/replacement – or its opposite, morpheme retention – and using the Swadesh list, he reports that approximately 81% of items are retained per millennium (Bynon, 1977; i.e. the rate of morpheme decay is 19% per millennium).[6] Lees (1953) has received considerable criticism on several counts; a large degree of arbitrary manipulation is needed in order to make the calculations work; the retention rate rests upon a sample of only 13 languages; these have a long written history – unlike most of the world's languages – which could introduce bias; and 11 of them are from the same (Indo-European) language family. Finally, the results of tests that seek to establish the validity of glottochronological methods are 'not encouraging' (Bynon, 1977) because they do not accord with known historical facts. For example, the split between French and Italian is placed in the sixteenth century, which is far too late.

The discussion so far has underlined the importance of, and difficulties associated with identifying cognates; the complexities of separating diachronic changes (those which happen over a long time) from borrowing (when items may be imported wholesale); and the uncertainty surrounding both the origin and meanings of Swadesh's classificatory scale. Glottochronology has also been briefly introduced and discussed. In Section 2, I turn to look at how lexicostatistical methods have been applied to sign language varieties and then, in Section 3, at the specific problems that have emerged in the process.

---

[5] This is problematic because not all linguists have understood this distinction. Consequently, those unfamiliar with lexicostatistical methods interpret lexicostatistical findings using terms such as 'genetic relationships' and 'genealogical relationships', in accordance with the more common definition of 'language family'.

[6] It is fascinating that Lees (1953) finds an 81% retention rate while Swadesh uses 81% as a threshold for classification. Both attempt to reflect a sense of how much change can happen before a certain level of divergence is reached, though Sheila Embleton doubts that there is an explicit link between them (p.c., 14 April 2013).

## 2. Lexicostatistical studies of sign language varieties

The first study to use lexicostatistical techniques also saw the first – and only known – attempt to apply glottochronology to sign languages. Woodward (1978) reports on the application of lexicostatistical methods from spoken language linguistics to American Sign Language (ASL) and Old French Sign Language (OFSL). This was motivated by the observation of similarities in the residual lexica of each sign language, alongside the knowledge that there was contact between the Paris National Institute for Deaf-Mutes and the deaf community in Hartford, via the American Asylum of the Deaf and Dumb, in the 1810s. Woodward applies glottochronological techniques in an attempt to show that sign languages change at a constant rate, on the basis of a comparison of diachronic change in ASL and Russian Sign Language (РЖЯ) – which is also reported to have split from OFSL (Wilbur, 1987).

Besides OLSF, ASL and РЖЯ, there are several other cases in the literature where sign languages are reported to be related. Typically in these cases, sign language users (deaf or hearing) have moved from one country to another, either temporarily or permanently. Through contact with deaf signers in the new country, sign language varieties from a migrant's country of origin are introduced, or to be more specific, lexical items are introduced, since it is not always clear whether grammatical structures are transmitted. Where a sign language variety already exists, it has been suggested that a process of creolisation takes place between the two sign languages (e.g. Woodward, 1978, in the case of OFSL and ASL). Other examples of such transmission are as follows:

**Geographically distant sign languages reported to be 'related'**
British Sign Language (BSL) to Auslan and New Zealand Sign Language (NZSL) (Schembri et al., 2010)
Japanese Sign Language (JSL) to South Korean Sign Language and Taiwan Sign Language (Su & Tai, 2009)
German Sign Language to Israeli Sign Language (Meir & Sandler, 2008)
Danish Sign Language to Icelandic Sign Language (Aldersson & McEntee-Atalianis, 2009).

In these cases, where a sign language variety has travelled hundreds or even thousands of miles, it is easy to see the motivation for using lexicostatistical methods to try to classify the extent to which language varieties have diverged. In some cases, it might be convincingly argued that these cases constitute true 'splits', despite the various practical and theoretical problems discussed in Section 1. Examples of this kind are perhaps more suited to the application of such lexicostatistical methods, where there are reasonable grounds to assume a discrete period of historic language contact, with little if any subsequent language contact. Unfortunately, there is not always much evidence to support such an assumption. For example, Aldersson and McEntee-Atalianis cite only the Ethnologue (Lewis, Simons and Fennig, 2014) as evidence for a historic relationship between Danish Sign Language and Icelandic Sign Language.[7]

Table 2 presents a selection of lexicostatistical studies that have been conducted on sign language varieties. It is striking that most of these lexicostatistical studies do not include known 'related' sign languages that have since split, but rather regional varieties within a specified geographical area – which is, in most cases, a single country. For example, Bickford (1989, 1991) does not include Mexican Sign Language (LSM) and another sign language that is geographically distant but known to be related; he includes several varieties from different regions of Mexico. Presumably the signers from at least some of these regions are in contact, whether regular or intermittent, and because of this, it is difficult if not impossible to show that formal similarities and potential cognates have not occurred through borrowing. Consequently, this kind of research is not concerned with historical relatedness.

---

[7] The Ethnologue (Lewis, Simons and Fennig, 2014) is not a reliable source for this kind of argument, as it contains many claims around the relationships between sign languages that are unsubstantiated by evidence.

**Table 2.** *A selection of studies that compare the closeness of sign language varieties using Swadesh's classificatory scale.*

| author (year) | sign language varieties |
|---|---|
| Woodward (1978) | ASL and OFSL |
| Bickford (1989, 1991) | varieties of Mexico |
| Woodward (1991) | varieties of Costa Rica |
| Woodward (1993) | varieties of India and Pakistan |
| Woodward (1996, 2000) | varieties of Thailand, Vietnam |
| McKee & Kennedy (2000) | BSL, Auslan, NZSL |
| Currie, Meier & Walters (2002) | JSL, ASL, French Sign Language (LSF), LSM |
| Hurlbut (2003) | varieties of Malaysia |
| Johnston (2003) | BSL, Auslan, NZSL |
| Bickford (2005) | varieties of Eastern Europe |
| Parkhurst & Parkhurst (2007) | varieties of Spain |
| Hurlbut (2008) | varieties of Taiwan |
| Hurlbut (2008) | varieties of Philippines |
| Johnson & Johnson (2008) | varieties of India |
| Parks and Parks (2008) | varieties of Guatemala |
| Hurlbut (2009) | varieties of Thailand |
| Aldersson & McEntee-Atalianis (2009) | Danish Sign Language and Icelandic Sign Language |
| Sasaki (2009) | JSL and Taiwan Sign Language (TSL) |
| Su and Tai (2009) | JSL, South Korean Sign Language, TSL and ASL |
| Parks and Parks (2010) | varieties of Peru |
| Padden (2011), Al Fityani & Padden (2010) | varieties of Jordan, Palestine, Kuwait and Libya |
| Hurlbut (2012) | varieties of Nepal |
| Isma (2012) | varieties of Java (Indonesia) |
| Hurlbut (2013) | varieties of Indonesia |

Most of the studies in Table 2 follow a similar approach: a word list is introduced, usually a modified Swadesh list, and data are obtained through elicitation, or dictionaries, or both. Most define the degree of lexical similarity between sign language varieties by categorising pairs of signs as 'identical', 'similar' or 'different' based on the number of corresponding phonological parameters such as handshape, orientation, movement and location – although this is not always straightforward, as Xu (2006) points out. The classificatory scale from lexicostatistics is then introduced.[8] On the back of this, different kinds of conclusions are reached. In some cases, the language varieties in question are labelled as dialects of the same language (Johnston, 2003; Hurlbut, 2008, 2013) or different languages (Isma, 2012). Alternatively, inferences are made concerning the mutual intelligibility of different varieties, or historical relatedness (Bickford, 1991).

Some of these studies are styled as rapid surveys of sign language varieties, and a typical example is Hurlbut (2013), written as a report entitled *The Signed Languages of Indonesia: An Enigma*.[9] On the basis of wordlist elicitation in an astounding 20 locations across the country, and subsequent comparison of responses in each location (Figure 2), Hurlbut (2013:18) applies lexicostatistical methods and concludes 'the results show clearly that Indonesian Sign Language is one language'.

---

[8] Many of the studies in Table 2.2 cite Crowley (1997) to justify their use of lexicostatistical methods, but none of these studies mention the practical and basic theoretical problems that Crowley describes (1997:183-186).

[9] This 20-page report became available at the end of 2013. I am grateful to Hope Hurlbut and Ted Bergman of SIL for allowing me to view a copy of it in advance.

Left matrix:

| | N Sumatra | W Sumatra | C Sumatra | S Sumatra | Jakarta (old) | Jakarta (new) | W Java | C Java | E Java | Yogyakarta | W Kalimantan | C Kalimantan | E Kalimantan | S Kalimantan | Gorontalo | N Sulawesi | C Sulawesi | S Sulawesi | Bali | NTT | Maluku |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **North Sumatra** | | | | | | | | | | | | | | | | | | | | | |
| **West Sumatra** | 69 | | | | | | | | | | | | | | | | | | | | |
| **Central Sumatra** | 80 | 80 | | | | | | | | | | | | | | | | | | | |
| **South Sumatra** | 73 | 68 | 80 | | | | | | | | | | | | | | | | | | |
| **Jakarta (old)** | 82 | 80 | 81 | 76 | | | | | | | | | | | | | | | | | |
| **Jakarta (new)** | 68 | 72 | 74 | 67 | 61 | | | | | | | | | | | | | | | | |
| **West Java** | 72 | 77 | 80 | 75 | 79 | 72 | | | | | | | | | | | | | | | |
| **Central Java** | 83 | 77 | 85 | 81 | 90 | 80 | 86 | | | | | | | | | | | | | | |
| **East Java** | 73 | 76 | 76 | 69 | 78 | 74 | 60 | 81 | | | | | | | | | | | | | |
| **Yogyakarta** | 77 | 64 | 74 | 72 | 72 | 59 | 70 | 81 | 69 | | | | | | | | | | | | |
| **West Kalimantan** | 67 | 79 | 78 | 73 | 76 | 60 | 69 | 70 | 64 | 61 | | | | | | | | | | | |
| **Central Kalimantan** | 68 | 81 | 80 | 72 | 81 | 70 | 73 | 78 | 75 | 64 | 74 | | | | | | | | | | |
| **East Kalimantan** | 71 | 77 | 80 | 78 | 79 | 70 | 78 | 76 | 74 | 68 | 76 | 76 | | | | | | | | | |
| **South Kalimantan** | 76 | 73 | 79 | 76 | 84 | 67 | 82 | 84 | 76 | 73 | 70 | 66 | 76 | | | | | | | | |
| **Gorontalo** | 66 | 81 | 81 | 72 | 82 | 71 | 75 | 78 | 79 | 62 | 72 | 81 | 75 | 72 | | | | | | | |
| **North Sulawesi** | 76 | 86 | 85 | 75 | 84 | 78 | 77 | 81 | 83 | 70 | 75 | 84 | 81 | 78 | 80 | | | | | | |
| **Central Sulawesi** | 65 | 83 | 85 | 70 | 77 | 66 | 71 | 74 | 75 | 60 | 71 | 82 | 73 | 82 | 82 | | | | | | |
| **South Sulawesi** | 70 | 65 | 74 | 77 | 67 | 57 | 70 | 68 | 63 | 70 | 63 | 62 | 69 | 65 | 70 | 68 | 67 | | | | |
| **Bali** | 81 | 72 | 77 | 74 | 77 | 71 | 73 | 83 | 76 | 75 | 63 | 69 | 75 | 80 | 68 | 76 | 70 | 70 | | | |
| **NTT** | 68 | 79 | 83 | 69 | 79 | 67 | 75 | 73 | 75 | 62 | 70 | 80 | 74 | 70 | 75 | 83 | 84 | 67 | 68 | | |
| **Maluku** | 68 | 83 | 83 | 72 | 81 | 69 | 70 | 77 | 77 | 67 | 79 | 88 | 75 | 73 | 76 | 90 | 84 | 65 | 70 | 83 | |
| **ASL** | 38 | 48 | 46 | 37 | 72 | 51 | 45 | 56 | 63 | 34 | 46 | 52 | 43 | 41 | 52 | 56 | 51 | 36 | 39 | 51 | 51 |

Right matrix (second stage):

| | Sumatra | Java | Kalimantan | Sulawesi |
|---|---|---|---|---|
| **Sumatra** | | | | |
| **Java** | 90 | | | |
| **Kalimantan** | 90 | 90 | | |
| **Sulawesi** | 91 | 90 | 90 | |
| **smaller islands** | 85 | 92 | 87 | 89 |

***Figure 2.*** *A comparison of the core lexicon of 20 urban sign communities across Indonesia (Hurlbut, 2013:18); and (top right) the second stage of Hurlbut's analysis (2013:19).*

Several preliminary linguistic studies have been conducted on sign language varieties in Jakarta and Yogyakarta. For Jakarta, this includes Chu and Wijaya (2013) on sign names; for Yogyakarta, Bharoto (2013) examines classifier constructions, and Sukmara (2014) explores phonological components. Lexicographical work has also commenced on both varieties (Woodward & Bharoto, 2011; Woodward, Wijaya & Satryawan, 2011), and again, lexicostatistical methods have been applied. These studies report that the Jakarta and Yogyakarta varieties share only 64% of their basic or core vocabulary. Accordingly,

> this percentage shows that Jakarta sign language and Yogyakarta sign language are not dialects of the same language, because for dialects from the same language, between 80 and 100% of basic vocabulary items are usually cognate (Woodward, Wijaya & Satryawan, 2011:vii, my translation).[10]

Isma (2012) presents this argument in its entirety, with a basic comparison of 100 items in Jakarta and Yogyakarta; she also compares the sign order of sentences that have reversible and irreversible arguments. These findings are based on a remarkably small sample of four signers from each city, and Isma (2012:73) notes that 'the sample size may not be representative enough'. When applied to Indonesian sign language varieties by Isma (2012) and Hurlbut (2013), lexicostatistical methods generate contradictory findings.

At this point, it is possible to conclude that most of the studies in Table 2 are preliminary lexicostatistical studies, because they do not follow classical lexicostatistics methods – historical relatedness is not established, and cognate status is not demonstrated. All of the logical and theoretical problems identified in Section 1 for spoken languages also apply to signed languages, but in Section 3, I deal with specific problems in applying lexicostatistics to *sign language* varieties.

## 3. Applying lexicostatistical methods to sign languages: Some problems

Sign linguists occasionally make critical notes on lexicostatistics (e.g. Woodward, 1991; Zeshan, 2000; Woll, Sutton-Spence & Elton, 2001; Meir & Sandler, 2008; Su & Tai, 2009) but these points have not been brought together into a single discussion. Furthermore, the dubious link that has emerged between lexicostatistical methods and sociolinguistic variation has not been clearly addressed in the sign language literature. In applying lexicostatistical methods, some make changes to try and mitigate particular problems (Woodward, 2000) while others add caveats that limit the intended scope of

---

[10] Exactly the same argument is presented in Woodward and Bharoto (2011).

investigation to synchronic study (McKee & Kennedy, 2000:54). However, lexicostatistical methods continue to be adopted, and I now explain my reservations regarding how these methods have been applied to sign language varieties. These concern the elicitation of items (3.1), iconicity (3.2), the word list (3.3), the 'cognate' (3.4) and the 'variation problem' (3.5).

## 3.1. Elicitation of items

Several studies collect data from dictionaries and similar publications, and Johnston (2003) gives an overview of the lexicographical issues that emerge when using this approach. Problems are also encountered when eliciting data directly from informants in the field. Eliciting specific lexical items requires pictures, or the knowledge of another written or signed language apart from the target language. In my own trials of lexicostatistical methods, use of cards with words written on them proved to be particularly disagreeable, with several informants seeming to find the process somewhat oppressive.[11] In some cases, the signed response replicated the morphological structure of the written stimulus. For example, when signing the Indonesian word *bermain* ('to play'), one participant used two signs, BER and MAIN (which incidentally is how *bermain* would be signed according to SIBI, the Indonesian Signed System).

The use of pictures is also not without problems: function words from the Swadesh list such as 'because' and 'if' cannot be rendered in pictorial form, and in any case these concepts might be expressed non-manually, rather than by a single sign. In addition, participants sometimes respond with descriptions of a picture, rather than individual lexical signs, suggesting that the visual form influences how the task is processed (Osugi, Supalla and Webb, 1999, and Nyst, 2007, have experienced similar problems with this method).[12]

## 3.2. Iconicity

Much has been written about the issue of iconicity (see Parkhurst & Parkhurst, 2003). Put succinctly, signs may be formally similar because they derive from the application of a similar metonymic process (Taub 2001:45), and not because they are historically related. Currie, Meier and Walters (2002:232) find that LSM and JSL are around 23% similar: historical relatedness and language contact cannot explain this finding, given that these language varieties are not known to be related, and have not been in extensive contact, which leaves only iconicity as a viable explanation. With this in mind, Woodward modified the Swadesh list to exclude signs that are likely to have an iconic basis, including pronouns and body parts – but there is no objective basis on which one can *predict* which signs are likely to be iconic. Bencie Woll (cited in Aldersson & McEntee-Atalianis, 2009:53) recommends including concepts with a propensity for iconic depiction 'because sign language users may produce these signs differently if they have different visual motivations'.[13]

## 3.3. The word list

The notion of a list of 'basic vocabulary' has itself come under fire from sign language linguists. McKee and Kennedy (2000) find it too restrictive, and question the accuracy of the findings of a comparison that uses Woodward's modified Swadesh list alone. They argue that a more random selection of lexemes should be used as the basis of a comparison, although Woodward (2011:40) has since criticised methods

---

[11] Sometimes, informants did not understand the meaning of the stimulus word, or showed signs of confusion, or provided an incorrect sign, for example confusing the Indonesian words *bintang* ('star') and *binatang* ('animal').
[12] Connie de Vos (personal communication, 25 April 2013) also reports that Kata Kolok signers sometimes provided recipes when asked for a lexical sign for a particular spice or dish.
[13] I recommend retaining all concepts *if the aim is to measure mutual intelligibility*, because if signers of different varieties produce the same sign, these will presumably be intelligible regardless of whether the sign is or is not iconic.

that use more than the basic core vocabulary. A further problem is semantic mismatch, known anisomorphism (see Brien & Turner, 1994): how can we know that the target sign language will have a sign for certain items on the wordlist? For example, the kin relations of 'brother' and 'sister' are lexicalised in English – the source language of the word list – but equivalent lexical items do not exist in any known sign language variety in Indonesia.

Isma (2012:23) notes that some words have more than one 'sense' and that there is no specification on how to resolve this. Furthermore, there is a cultural mismatch, and some terms on Swadesh's word list – such as 'louse' and 'grease' – are not closely associated with sign communities (Woll, Sutton-Spence & Elton, 2001). 'Snow', which is on the basic vocabulary list, is in no sense part of the 'basic vocabulary' of most signers living in tropical areas, and it is perhaps more likely that a signer will create an idiolectal sign for this, in the absence of a more conventional form.[14] With these difficulties in mind, it makes sense to approach language documentation from the perspective of the target language, avoiding anisomorphism by documenting signs as they appear in context (incidentally, this is one of the advantages of using spontaneous corpus data rather than lexical elicitation).

## 3.4. Historical linguistics and the 'cognate'

In Section 1, I explained that two forms are cognate if they derive from the same source; that is, from the same proto-item in a common parent language. Lexicostatistical studies of sign languages face the problem of how to identify cognate status. Lack of written documentation, needed in order to chart the history of a language, is a problem faced by many spoken language linguists, as well as sign linguists. However, it is possible to reconstruct proto-languages through the identification of regular sound correspondence, because specific and regular patterns between certain forms are unlikely to occur by chance alone (Crowley, 1997). It is now well-established that sign languages have a structure at the phonological level, but only a small number of diachronic changes have been identified at this level (Frishberg, 1975), and it is not clear how far these changes are akin to regular sound correspondence in spoken languages.

Based on a comparison of ASL and OFSL, Frishberg (1975) lists changes that seem to have occurred to signs over time: iconic gestures become more arbitrary, signs become displaced in certain ways within the signing space, one-handed signs become two-handed, and non-symmetrical signs become symmetrical. However, it is difficult to see how these processes could help to infer cognate status in cases where there is no other reason to suppose a historical relationship between varieties, or between forms. For example, nowhere is it attested that Variety A has a series of one-handed signs, that Variety B has a corresponding set of two-handed signs, and hence that Variety A and Variety B are related. This suggests that the changes Frishberg identifies are not truly equivalent to diachronic phonological changes in spoken languages.

Woodward (2011:41) helpfully suggests some of the possible phonological rules that may be responsible for deriving a current form from an earlier one, including assimilation, dissimilation, deletion, epenthesis, coalescence and metathesis, but he does not explain how these processes can reliably reveal proto-forms when only the current forms are known. Notably, in none of his work does Woodward describe the basis on which he attributes cognate status to forms in sign languages where historic documentation is unavailable. Some sign linguists look for these changes taking place between the varieties used by older and younger signers, but this brings us no closer to identifying cognate pairs across sign language varieties.

There are two other associated problems. First, historical linguists stipulate that, since cognate status precludes borrowing phenomena, great care must be taken to remove borrowed items from cognate

---

[14] These difficulties are of course not restricted to sign languages: see for example Huang et al. (2007).

counts. There is usually no way of knowing how to differentiate between borrowed and historically derived forms (Embleton, 2000; Meir & Sandler, 2008; Lanesman, 2013). Particularly in cases where the language varieties in question are regional and proximate, similarities due to areal contact is highly likely, and in Section 2 it was shown that many studies seeking to apply lexicostatistical methods focus on such proximate varieties. Indeed, it is natural for users of proximate language varieties to borrow from each other, and the lexicon is known to be the easiest element of a language to borrow (Muysken, 1995). It is therefore necessary to be cautious when drawing conclusions about the implications of lexical similarity.

The second point concerns the Stammbaum, or 'family-tree' model, which has been used by historical linguists for many years to describe the way in which languages change and split into different languages through time. Some sign linguists have applied the concept of the language family directly to sign languages, and this is problematic because the notion of language families and genetic relationships is not well-defined for sign language research (Palfreyman, Sagara & Zeshan, 2015). For example, many of the premises of historical linguistics are based on the documented history of the Indo-European language family over several centuries. In the case of Indonesia, for example, we do not know for certain the time depth of its sign language varieties, but it seems highly unlikely that they have been used continuously for several centuries.

Even in cases where there is evidence that sign language was used in the distant past, we do not know if the 'sign language' has been used continuously.[15] In other words, the time scale of sign languages is likely to be very different from many spoken languages that have a long and unbroken history (Woll, Sutton-Spence & Elton, 2001:22). Considerable caution is therefore needed when applying concepts from historic linguistics to sign languages.[16] Remember, too , that where spoken languages *are* known to have long and unbroken histories, spoken language linguists (e.g. Embleton, 2000) question the appropriateness of the 'family-tree model', and wonder what exactly is meant by 'a group of genetically related languages'.

## 3.5.  The variation problem

The majority of lexicostatistical studies make no mention of the possibility that signers themselves might know and use more than one variant, and from a sociolinguistic perspective this is perhaps the method's most notable shortcoming. Even when studies acknowledge and investigate variation within one region, most of them only seem to elicit one variant from each participant, or at any rate, they make no mention of how they deal with the (highly likely) possibility that signers may know more than one variant.[17] A notable exception is Stamp, who states that, as part of her investigation of colour and other concepts:

> participants were asked [...] to produce any other signs they knew for that concept (e.g.,
> regional, informal/formal variants). The first sign produced was considered to be the

---

[15] For example, Miles (2000) shows that sign language was used in the court of the Ottoman Empire, but unfortunately there is no evidence to suggest that this historic variety is related to modern Turkish Sign Language. To paraphrase Dixon (1997:37f), for questions concerning the time depth and development of most sign languages, there is only one honest answer: 'we don't know'.

[16] Woodward (1978) suggests that sign languages evolve at the same rate as spoken languages, based on research of the kind conducted by Frishberg (1975), but as discussed earlier, there is no compelling evidence for this.

[17] Hurlbut (2013) is the first lexicostatistical study I have seen that mentions this issue. She notes that: 'a working assumption was that everyone living in the same city would know all the signs used by those from whom I was eliciting the words. Thus when comparing two cities with each other if a sign from one subject was the same or similar to any one of the signs from another city, the two were counted as similar for that item' (Hurlbut, 2013:10). There are still problems: in each place, data was elicited from 'one or more' signers (the number varied); where data was elicited from several signers, Hurlbut concedes that social networks can have a decisive impact on the outcome of comparison (p.14).

signer's default variant, unless the signer stated explicitly that another variant was the one they use most on a daily basis (Stamp, 2013:142).

Notably, however, Stamp's research is not lexicostatistical. I suspect that most if not all *lexicostatistical* studies avoid the question of whether signers have an active or passive knowledge of more than one variant because the lexicostatistical method cannot cope with this possibility.[18]

Originally, the question of variation was not significant because the method was only intended to examine the number of items across two historically-related language varieties for which potential cognates exist, with a view to estimating the point at which those varieties split. But since the method has been applied to the sole task of delineating language varieties, it no longer makes sense to ignore intra-varietal or intra-individual variation. The apparent working assumptions of the lexicostatistical method – that language varieties are homogeneous and that the signs used by individuals do not vary – are without foundation.

Consider the following hypothesis. A word list is used to elicit data from two signers who live in the same city on separate occasions. When asked what sign they use for an item on the word list, Signer A produces variant *x*, and Signer B produces variant *y*. As signs *x* and *y* are formally different, they are not 'cognate', which lowers the overall percentage of cognates. If this effect occurs often enough, and the overall percentage happens to fall below 80%, the language varieties of each signer are classified as 'different languages'. Yet if we are able to ascertain that Signer A and Signer B both know (and perhaps use) signs *x* and *y*, the resulting conclusion of 'different languages' is highly inaccurate.[19]

## 4. Comments on the application of lexicostatistical methods

It seems that, when applying lexicostatistical methods, many sign language linguists have actually been less interested in the historical relatedness of languages, and more interested in mutual intelligibility and in delineating languages along the language-dialect continuum (Parkhurst & Parkhurst, 2003). Notably, many of the studies in Table 2 do not discuss relationships between sign languages lower down on Swadesh's classificatory scale: sign language varieties are not grouped as belonging to the same 'stock' or 'mesophylum' in the way that some spoken languages are. In basing methods on lexicostatistical principles, the result is a mismatch between research design and research aim (Figure 3). Once it is accepted that the actual aim of a study is to examine mutual intelligibility, the premises and assumptions underlying the method must be re-examined, since it is not appropriate to transfer methods intended for one research question to answer a different research question.

The implicit aim – of measuring mutual intelligibility – has important implications for questions concerning how to solve some of the problems that have been identified. Most of these problems have been tied in some way to lexicostatistical concerns, but if the actual purpose is to determine mutual intelligibility, we need be confined by lexicostatistical theories no longer, and different questions must be asked instead. For example, linguists have attended to the question of how to exclude iconic signs in order to avoid treating as cognate signs that are actually unrelated, and similar only due to the same iconic motivation. Once the aim is to measure mutual intelligibility, it is surely important to *include* consideration of iconic signs.

---

[18] This is the likely explanation for the opposite conclusions drawn by Isma (2012), Woodward, Wijaya and Satryawan (2011) and Woodward and Bharoto (2011) on one hand, and Hurlbut (2013) on the other, even though both use 'lexicostatistical methods'.

[19] Even if Signer A were to use a variant form that Signer B did not know, it seems unlikely that this would cause a major misunderstanding; Signer B might use contextual information, or mouthings, or interpret the iconic properties of the variant, to understand what Signer A is saying.
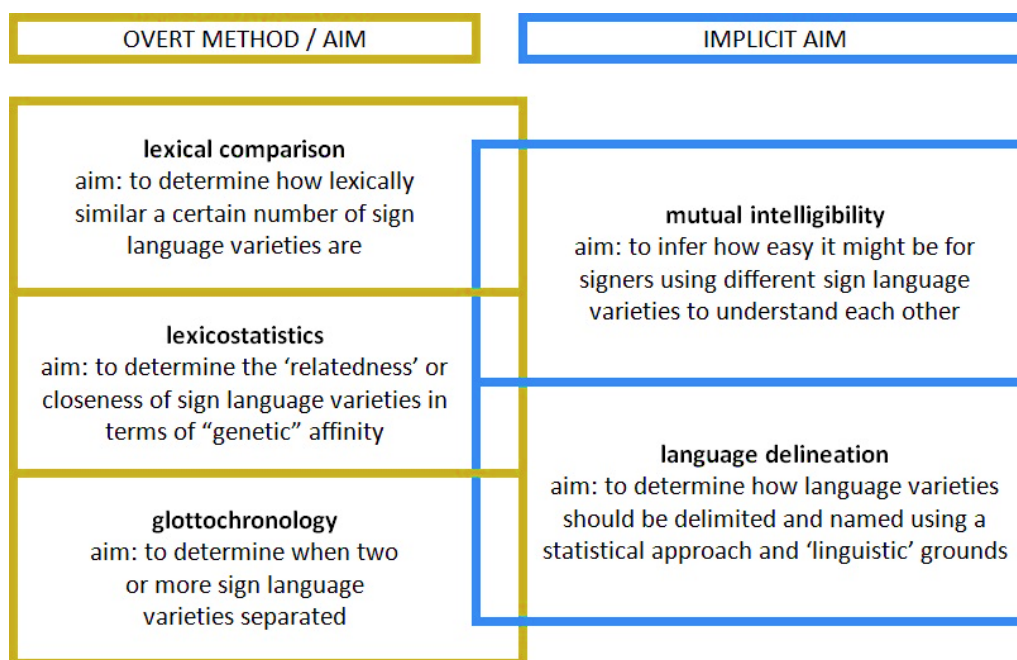
**Figure 3.** *Overt methods and implicit aims of lexicostatistical research.*

Another premise requiring re-evaluation is the decision to compare only a section of the lexicon. As explained above (Section 1), this decision was originally based on the assumption that the 'core vocabulary' is more resistant to change. However invalid this assumption might be, once the overt goal is mutual intelligibility and not lexicostatistics, it is no longer necessary to limit our focus to a small portion of the lexicon, and indeed, the prospect of linguistic comparison at other levels of organisation becomes viable, including the consideration of phonology, morphosyntax and semantics.

It is not hard to see why linguists have used lexicostatistical methods. The idea of quantifying mutual intelligibility adds a perceived measure of 'objectivity' to the complex issue of language delineation, while the mathematical scale and cosmetic concerns over removing 'iconic' signs from the word list lend an additional air of credibility. The use of a short word list also enables quicker results than other methods, such as the collection, annotation and analysis of the same (or more) lexical items in a corpus of natural data, or the analysis of grammatical structure. However, the results of studies such as Isma (2012) and Hurlbut (2013) are necessarily partial, and have the potential to confuse sign community members who may not understand either the method or the outcomes.

The suitability of Swadesh's classificatory scale for measuring mutual intelligibility has not, to my knowledge, been discussed in the literature on spoken language linguistics, least of all sign linguistics. According to Crowley (1997:185), 'it seems that as soon as speakers of two different speech traditions... have more than about a 20 per cent difference in their basic lexicons, then mutual intelligibility is lost'. Yet from what I can establish, this assertion has no substantial basis: Swadesh's scale was entirely arbitrary (see Section 1), and even if the 81% threshold could be applied successfully to all spoken language varieties, the usefulness of this scale for the mutual intelligibility of signed language varieties remains unclear.

The arbitrary nature of any scale that seeks to delineate varieties along the language-dialect continuum is apparent from Isma (2012:33), where it is concluded that 79.7% of signs are 'cognate' – this is 1.3% short of classifying the varieties of Jakarta and Yogyakarta as dialects of the same language. Furthermore, Zeshan (2000) and Hendriks (2007) suggest that the outcomes of lexicostatistical studies do not necessarily correspond with mutual intelligibility: signers have strategies for dealing with variation, and are perhaps more experienced in dealing with lexical variation than are speakers.

## 5. The need to move away from lexicostatistical approaches

Given the various practical and theoretical problems associated with lexicostatistical methods, such methods are not viable for analysing variation in sign languages or for delineating sign language varieties. Most of the studies that seek to apply lexicostatistical methods fall short of the requirements of classical lexicostatistics, in part because of the need to show beyond doubt that lexical similarities are due to historical relatedness and not borrowing or iconicity.

As a means of quantifying delineation along the language-dialect continuum, the lexicostatistical approach falls short, most notably in its failure to account for the existence of several variants. Lexicostatistics has been used to quantify variation, but paradoxically it cannot deal with it.

This is not to imply that lexical comparison is without value. Clearly, the question of whether the lexica of two varieties have many, few or no overlaps is of great significance. However, there is little point in adopting a research design based on lexicostatistical methods in a bid to delineate sign languages, and any study must find alternative and valid ways of addressing questions concerning iconicity, variation, and mutual intelligibility.

Studies that aim to delineate sign language varieties have often surmised that the quantification of linguistic similarities and differences is a suitable proxy for mutual intelligibility, usually focusing on lexical comparison. With some notable exceptions, those studies have applied and misconstrued lexicostatistical methods created in the 1950s for spoken language varieties, while disregarding sociolinguistic variation and – perhaps most importantly – the perspectives of signers themselves. If the field is to move forwards, sign language linguists must set lexicostatistics aside and seek new ways of dealing with the language delineation question.

## References

Aldersson, Russell and Lisa McEntee-Atalianis
2009    A lexical comparison of signs from Icelandic and Danish Sign Languages. *Sign Language Studies,* 9:1, 45-87.

Al-Fityani, Kinda and Carol Padden
2010    Sign languages in the Arab world. In: Diane Brentari (ed.) *Sign Languages: A Cambridge Language Survey.* Cambridge: Cambridge University Press, 433-450.

Arndt, Walter W.
1959    The performance of glottochronology in German. *Language* 35, 180-92.

Bergslund, Knut and Hans Vogt
1962    On the validity of glottochronology. *Current Anthropology* 3, 115-58.

Bharoto, Adhi Kusuma
2013    Classifier constructions in the Yogyakarta variety of Indonesian Sign Language. A paper presented at the Third International Conference on Sign Linguistics and Deaf Education in Asia, CUHK, 31 January – 2 February 2013.

Bickford, Albert J.
1989    *Lexical Variation in Mexican Sign Language.* Work Papers of the Summer Institute of Linguistics. Grand Forks: ND.
1991    Lexical Variation in Mexican Sign Language, *Sign Language Studies* 72, 241-276.
2005    The signed languages of Eastern Europe. SIL Electronic Survey Reports.

Brien, David and Graham Turner
1994    Lemmas, dilemmas and lexicographical anisomorphism: Presenting meanings in the first BSL-English dictionary In: Inger Ahlgren, Brita Bergman and Mary Brennan (eds.), *Perspectives on Sign Language Structure - Papers from the Fifth International Symposium on Sign Language Research* (Vol. 2). Durham: The International Sign Linguistics Association, 391-407.

Bynon, Theodora
  1977    Historical Linguistics. Cambridge: Cambridge University Press.

Campbell, Lyle
  1977    Quichean linguistic prehistory. Berkeley, CA: University of California.
  2004    *Historical Linguistics: An Introduction* (second edition). Edinburgh: Edinburgh University Press.

Chu, Kenny and Laura Lesmana Wijaya
  2013    The linguistic and sociocultural aspects of name signs: a comparison between Hong Kong Sign Language (HKSL) and Indonesian Sign Languages (Jakarta). A paper presented at the Third International Conference on Sign Linguistics and Deaf Education in Asia, CUHK, 31 January – 2 February 2013.

Crowley, Terry
  1997    *An Introduction to Historical Linguistics* (third edition). Oxford: Oxford University Press.

Currie, Anne-Marie Guerre, Richard P. Meier and Keith Walters
  2002    A cross-linguistic examination of the lexicons of four signed languages. In: Richard P. Meier, Kearsy Cormier and David Quinto-Pozos (eds.) *Modality and Structure in Signed and Spoken Languages*, Cambridge: Cambridge University Press., 224-236.

Dixon, Robert, M. W.
  1997    *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Embleton, Sheila M.
  1992    Historical linguistics: Mathematical concepts. In: William Bright (ed.) International Encyclopedia of Linguistics, Volume 2. Oxford: Oxford University Press. 131-5.
  2000    Lexicostatistics/Glottochronology: From Swadesh to Sankoff to Starostin to future horizons. In: Colin Renfrew, April McMahon and Larry Trask (eds.), *Time Depth in Historical Linguistics.* Cambridge: McDonald Institute for Archaeological Research, 143-165.

Frishberg, Nancy
  1975    Arbitrariness and iconicity: Historical change in ASL. *Language* 51, 696–719.

Gudschinsky, Sarah, C.
  1956a   The ABCs of lexicostatistics (glottochronology). In: Dell Hymes (ed.), *Language in Culture and Society: A Reader in Linguistics and Anthropology*. New York: Harper and Row, 612-623.
  1956b   Three disturbing questions concerning lexicostatistics. *International Journal of American Linguistics* 22:3, 212-213.

Hammarström, Harald, Robert Forkel, Martin Haspelmath and Sebastian Nordhoff
  2014    Glottolog 2.3, Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://glottolog.org, retrieved 12 June 2014.

Hawkins, John A.
  2009    Germanic languages. In: Bernard Comrie (ed.) *The World's Major Languages* (second edition). London: Routledge, 51-58.

Hendriks, Bernadet
  2007    Negation in Jordanian Sign Language: A cross-linguistic perspective. In: Perniss, Pfau and Steinbach, *Visible Variation: Comparative Studies on Sign Language Structure.* Berlin: de Gruyter, 103-128.

Hoijer, Harry
  1956    Lexicostatistics: A critique. *Language* 32, 49-60.

Huang, Chu-Ren, Laurent Prevot, I-Li Su and Jia-Fei Hong
  2007    Towards a conceptual core for multicultural processing: A multilingual ontology based on the Swadesh list. *Lecture Notes in Computer Science,* 4568, 17-30.

Hurlbut, Hope M.
  2003    A preliminary survey of the signed languages of Malaysia. In: Anne Baker, Beppe van den Bogaerde and Onno Crasborn (eds.) *Cross-linguistic perspectives in sign language research: Selected papers from TISLR 2000*. Hamburg: Signum Verlag, 31-46.

2008a     *Philippine signed languages survey: A rapid appraisal.* SIL International

2008b     *A survey of sign language in Taiwan*. SIL International.

2009      *Thai signed languages survey: A rapid appraisal*. SIL International.

2013      *The signed languages of Indonesia: An enigma*. SIL International.

Isma, Silva Tenrisara Pertiwi

    2012      Signing varieties in Jakarta and Yogyakarta: Dialects or separate languages? MA thesis, CUHK, Hong Kong.

Johnston, Trevor

    2003      BSL, Auslan and NZSL: Three signed languages or One? In: Anne Baker, Beppe van den Bogaerde and Onno Crasborn (eds.) *Cross-linguistic perspectives in sign language research: Selected papers from TISLR 2000*. Hamburg: Signum Verlag, 47-70.

Joseph, Brian D. and Richard D. Janda

    2007      *The Handbook of Historical Linguistics.* Wiley-Blackwell: Oxford.

Lanesman, Sara

    2013      *Algerian Jewish Sign Language: Its emergence and survival.* MA thesis, University of Central Lancashire, Preston.

Lees, Robert B.

    1953      The basis of glottochronology. *Language* 29:2, 113-127.


Lewis, M. Paul, Gary F. Simons and Charles D. Fennig (eds.)

    2014      *Ethnologue: Languages of the World.* Seventeenth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com

McKee, David and George Kennedy

    2000      Lexical comparisons of signs from American, Australian, British and New Zealand Sign Languages. In: Karen Emmorey and Harlan Lane (eds.), *The Signs of Language Revisited: An Anthology to Ursula Bellugi and Edward Klima.* Mahwah, NJ: Erlbaum, 49-76.

Meir, Irit and Wendy Sandler

    2008      *A Language in Space: The Story of Israeli Sign Language.* New York: Lawrence Erlbaum Associates.

Muysken, Pieter

    1995      Code-switching and grammatical theory. In: Lesley Milroy and Pieter Muysken (eds.) *One Speaker, Two Languages: Cross Disciplinary Perspectives on Code-Switching*. Cambridge: Cambridge University Press.

Nyst, Victoria

    2007      A descriptive analysis of Adamorobe Sign Language (Ghana). PhD dissertation, University of Amsterdam.

Osugi, Yutaka, Ted Supalla and Rebecca Webb

    1999      The use of word elicitation to identify distinctive gestural systems on Amami Island. *Sign Language and Linguistics*, 2, 87-112.

Padden, Carol

    2011      Sign language geography. In: Gaurav Mathur and Donna Jo Napoli (eds.), *Deaf Around the World: The Impact of Language.* Oxford: Oxford University Press, 19-37.

Palfreyman, Nick, Keiko Sagara and Ulrike Zeshan

   in press    Methods in carrying out language typological research. To appear in: Eleni Orfanidou, Bencie Woll and Gary Morgan (eds.). *Research Methods in Sign Language Studies: A Practical Guide*. Oxford: Wiley-Blackwell.

Parkhurst, Stephen and Dianne Parkhurst

    2001      Variación de las Lenguas de Signos Usadas en España: Un Estudio Lingüístico. Revista Española de Lingüística de las Lenguas de Signos.

    2003      Lexical comparisons of signed languages and the effects of iconicity. Work Papers of the Summer Institute of Linguistics, University of North Dakota Session, Volume 47.

Parks, Elizabeth and Jason Parks

2008      Sociolinguistic Survey Report of the Deaf Community of Guatemala. *SIL Electronic Survey Report* 2008-016.

2010      A Sociolinguistic Profile of the Peruvian Deaf Community. *Sign Language Studies*. 10:4, 409-441.

Schembri, Adam, Kearsy Cormier, Trevor Johnston, David McKee, Rachel McKee and Bencie Woll.

2010      Sociolinguistic variation in British, Australian and New Zealand Sign Languages. In: Diane Brentari (ed.) *Sign Languages: A Cambridge Language Survey.* Cambridge: Cambridge University Press, 476-498.

Stamp, Rose

2013      Sociolinguistic variation, language change and dialect contact in the British Sign Language (BSL) lexicon. PhD dissertation, UCL, London.

Starostin, George

2010      Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3, 79-116.

Su, Shiou-fen and James H-Y. Tai

2009      Lexical comparison of signs from Taiwan, Chinese, Japanese, and American Sign Languages: Taking iconicity into account. In: James H-Y. Tai and Jane Tsay (eds.) *Taiwan Sign Language and Beyond.* Chia-Yi, Taiwan: National Chung Cheng University, 149-176.

Swadesh, Morris

1950      Salish internal relationships. *International Journal of American Linguistics,* 16, 157–167.

1954      Perspectives and problems of Amerindian comparative linguistics, *Word,* 10, 306-332.

1955      Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics,* 21, 121-137.

Taub, Sarah F.

2001      *Language from the Body: Iconicity and Metaphor in ASL.* Cambridge: Cambridge University Press.

Teeter, Karl V.

1963      Lexicostatistics and genetic relationship. *Language,* 39, 638-48

Trask, Robert L.

1996      *Historical Linguistics.* London: Arnold.

Warnow, Tandy

1997      Mathematical approaches to computational linguistics. *Proceedings of the National Academy of Sciences of the United States of America.* 94(13), 6585-6590.

Woll, Bencie, Rachel Sutton-Spence and Frances Elton

2001      Multilingualism: The global approach to sign languages. In: Ceil Lucas (ed.) *The Sociolinguistics of Sign Languages.* Cambridge: Cambridge University Press, 8-32.

Woodward, James

1972      Implications for sociolinguistic research among the Deaf. *Sign Language Studies,* 1.

1978      Historical bases of American Sign Language. In: Patricia Siple (ed.) *Understanding Language through Sign Language Research.* New York: Academic Press.

1991      Sign language varieties in Costa Rica, *Sign Language Studies,* 73, 329-346.

1993      The Relationship of Sign Language Varieties in India, Pakistan, and Nepal. *Sign Language Studies,* 78, 15-22.

1996      Modern Standard Thai Sign Language, influenced from ASL, and its relationship to Original Thai Sign Language varieties. *Sign Language Studies,* 92, 227-252.

2000      Sign languages and sign language families in Thailand and Vietnam. In: Karen Emmorey and Harlan Lane (eds.), *The Signs of Language Revisited: An Anthology to Ursula Bellugi and Edward Klima.* Mahwah, NJ: Erlbaum, 23-47.

2003      Sign languages and Deaf identities in Thailand and Viet Nam. In L. Monaghan, C. Schmaling, K. Nakamura and G. Turner (eds.), *Many Ways to Be Deaf: International Variation in Deaf Communities.* Washington, DC: Gallaudet University Press.

    2011       Some observations on research methodology in lexicostatistical studies of sign languages. In: Gaurav Mathur and Donna Jo Napoli (eds.), *Deaf Around the World: The Impact of Language.* Oxford: Oxford University Press, 38-53.

Woodward, James and Adhi Kusuma Bharoto
    2011       Yogyakarta Sign Language dictionary. Jakarta: University of Indonesia Press.

Woodward, James, Laura Lesmana Wijaya and Iwan Satryawan
    2011       Jakarta Sign Language dictionary. Jakarta: University of Indonesia Press.

Xu, Wang
    2006       A comparison of Chinese and Taiwan Sign Languages: Towards a new model for Sign Language Comparison. MA thesis, Ohio State University.

Zeshan, Ulrike
    2000       *Gebardensprachen des indischen Subkontinents* [Sign languages of the Indian Subcontinent]. PhD dissertation, LINCOM Europa, Muenchen.