

# Building BSL SignBank: The lemma dilemma revisited<sup>1</sup>

## Abstract

One key criterion when creating a representation of the lexicon of any language within a dictionary or lexical database is that it must be decided which groups of idiosyncratic and systematically modified variants together form a lexeme. Few researchers have, however, attempted to outline such principles as they might apply to sign languages. As a consequence, some sign language dictionaries and lexical databases appear to be mixed collections of phonetic, phonological, morphological, and lexical variants of lexical signs (e.g. Brien 1992) which have not addressed what may be termed as *the lemma dilemma*. In this paper, we outline the lemmatisation practices used in the creation of BSL SignBank (Fenlon et al. 2014a), a lexical database and dictionary of British Sign Language based on signs identified within the British Sign Language Corpus (<http://www.bsllcorpusproject.org>). We argue that the principles outlined here should be considered in the creation of any sign language lexical database and ultimately any sign language dictionary and reference grammar.

Keywords: lemma, lexeme, lemmatisation, sign language, dictionary, lexical database.

## 1 Introduction

When one begins to document the lexicon of a language, it is necessary to establish what one considers to be a lexeme. Generally speaking, a lexeme can be defined as a unit that refers to a set of words in a language that bear a relation to one another in form and meaning. These words can be represented by a headword or lemma, often the canonical form from which other variants appear to be derived. For instance, in English, **walk**, **walks**, **walking**, and **walked** are all considered to be variant forms of the lemma **walk**. Spoken language dictionaries (e.g. Collins English Dictionary and the Oxford English Dictionary) have widely adopted the practice of listing words primarily by lemmas (e.g. **walk**, **walks**, **walking**, etc. are not listed separately in the Oxford English Dictionary but under **walk**). In principle, dictionaries of signed languages need not be organised any differently. However, before one can begin to list the signs in a sign language, one needs to develop clear criteria defining what are considered to be variant forms of the same lexeme and what might be considered separate lexemes (e.g. Johnston and Schembri 1999). Unfortunately, few sign language dictionaries appear to organise their entries in this manner, so lemmatisation (i.e. identifying variant forms of a lexeme and assigning a headword to represent the lexeme) for the purposes of lexical database building may not yet have been carried out for many sign languages.

This task is complicated by the need to group together phonetic and phonological variants without reference to a standard writing system.

This paper intends to address this issue in relation to the creation of BSL SignBank (Fenlon et al. 2014a), a lexical database and dictionary of British Sign Language (BSL) based on signs identified within the British Sign Language Corpus (Schembri et al. 2014). As no appropriate lemmatised resources for BSL existed prior to the creation of BSL SignBank, signs found in the BSL Corpus needed to be lemmatised in a principled way so that they could be added to BSL SignBank. This was done in part building on the work of Johnston and Schembri (e.g. Johnston 1989; Johnston and Schembri 1999) for Australian Sign Language, which is historically related to BSL. Here we examine in closer detail the procedures that need to be followed in order to lemmatise sign language data. In Section 2, we briefly describe the notion of lemmatisation in relation to sign languages and discuss previous attempts to develop lemmatisation principles for sign languages. Section 3 provides a brief overview of BSL SignBank and its association with the BSL Corpus. In Section 4, we outline the lemmatisation principles used in the creation of BSL SignBank building on previous work with other sign languages and enhancing our description with problematic cases that arose during our efforts. In Section 5, we argue that the principles outlined here should be considered in the creation of any sign language lexical database and discuss further issues involved in the development of sign language dictionaries.

## **2 Sign language lexicography**

### *2.1 Challenges in sign language lexicography*

Sign language research is still in its infancy compared to research on spoken languages. It was not until the mid twentieth century that linguistic descriptions of sign languages emerged (Tervoort 1953; Stokoe 1960). Since then, sign languages have attracted the attention of a growing number of researchers working across the language sciences and investigations of many aspects of sign languages as living languages have emerged (see Pfau et al. 2012 for an extensive review). Despite major advances in our understanding of sign languages, however, there remains much to be explored. Many of the world's sign languages lack a dictionary that is intended to be representative of the core lexicon of that language. There are good reasons for the absence of such a resource. Firstly, there is no standard orthography or notation system for sign languages. Secondly, sign language dictionaries have been (until very recently) unable to exploit large corpora to obtain a collection of signs for inclusion in dictionaries. Instead, word lists are often created simply by drawing on the spoken language of the surrounding community, listing key concepts from that language, and providing translation equivalents in the corresponding sign language. This practice

has been widely criticised by sign language lexicographers as such an approach presents a misleading and limited view of the sign language lexicon (e.g. Brien and Turner 1994; Johnston and Schembri 1999; Zwitserlood 2010). Not only does this practice imply that there is a one-to-one correspondence between spoken language words and signs, it is likely that the resulting collection of signs will not be very representative of the sign language lexicon. The issue of how to order signs within a printed dictionary without a standard writing system is also problematic; some lexicographers advocate the practice of listing signs according to formational criteria specific to sign languages (e.g. handshape) but they vary in the way they do this. This was a practice first adopted by Stokoe et al. (1965) for American Sign Language and replicated in, for example, the Auslan (Australian Sign Language) dictionary (Johnston 1989) and the Dictionary of British Sign Language/English (Brien 1992). Lastly, sign language dictionaries are inherently bilingual or multilingual dictionaries (we are not aware of any monolingual sign language dictionaries). Whilst signs may be presented in pictures, translation equivalents and definitions are presented in a written language (e.g. the language of the surrounding hearing community) and the opportunity to search for a sign via the spoken language may be a preferred option for many. Many sign language lexicographers have suggested ways to overcome these issues. In fact, with the advent of electronic dictionaries and sign language corpora, some of the issues may no longer need to hinder the development of sign language dictionaries (Kristoffersen and Troelsgård 2012), though there are still challenges to overcome (Zwitserlood 2010; McKee and McKee 2013). In this paper, we wish to focus on a specific problem that all sign language lexicographers must contend with: lemmatisation. This issue is important now that sign language corpora are becoming more widespread, since these corpora require lemmatised resources to assist in the process of annotation.

## 2.2 *Lemmatisation and sign languages*

The traditional notion of lemmatisation for spoken languages refers to the practice of grouping together morphological variants under a single lemma. For example, in English, the variants **jump**, **jumps**, **jumping**, and **jumped** are all considered variants of a single lexeme that is represented by the lemma **jump** in a dictionary. Generally speaking, lexicographers tend to follow the division between inflectional and derivational morphology as a guide when distinguishing lexemes (Sterkenburg 2003). Inflectional variants (which arise out of a process that modifies a word) are not recognised as separate lexemes whilst derivational variants (arising out of a process that forms new words) are treated as such. Following this, **jumps**, **jumped**, and **jumping** are forms of a single lexeme since each variant is derived by the addition of an inflectional suffix marking tense or aspect. In contrast, **teach** and **teacher** would be recognised as two separate lexemes

because **teacher** is formed through the addition of an agentive suffix (i.e. a derivational morpheme) that changes the part of speech of the word from a verb to a noun. The change in meaning that results from the latter process is different to the former; whilst the addition of a derivational morpheme changes the overall meaning and/or the word class of a lexeme, the addition of an inflectional morpheme only augments meaning and has a grammatical function (e.g. marking tense, number, agreement, etc.).

However, developing consistent lemmatisation principles along the same lines for sign languages can be difficult. Whilst some researchers argue that the division of inflectional and derivational morphology can be applied to sign languages (e.g. Sandler and Lillo-Martin 2006), others have argued otherwise. For example, Liddell (2003) has claimed that American Sign Language (ASL) is basically an inflectionless language, as there is no tense marking, aspectual modifications and plural marking are optional, and the so-called verb agreement system is perhaps best analysed as a reference tracking system. Liddell instead proposes that signs modified for aspect, plurality and person-marking are better understood as part of a derivational system. Similar claims have been made for other sign languages (e.g. Bergman and Dahl 1994; Johnston and Schembri, 2007). The categorisation of modified forms into inflectional and derivational variants is further complicated by the fact that it can be difficult to determine the part of speech of a sign (e.g. see Section 4.3.1) and since similar modifications may be regarded as inflectional in one instance but derivational in another (e.g. compare the aspectual modification in WALK to the movement seen in ALCOHOLIC described in Section 4.2.1). The ambiguity between the two categories and a lack of general consensus in the literature causes problems for the sign language lexicographer; it is not always clear how principles of lemmatisation for sign languages can be derived using structural distinctions borrowed from spoken languages like English.

As noted in Section 2.1, the issue of lemmatisation is complicated further as there are no standard writing systems widely used for any sign language. For spoken languages with written orthographies, the lexicographer is able to ignore phonetic and phonological variants, which makes distinctions between potentially related lexical variants straightforward.<sup>2</sup> For example, in English, variable pronunciations of **bath** such as [bæθ] and [baθ] would be paired together automatically as they both are represented identically in standard English orthography. With sign languages, variation can also exist at the phonetic and phonological level but without a standard writing system, the sign language lexicographer must also contend with the task of grouping together phonetic and phonological variants in a principled way.<sup>3</sup> One can therefore see that what must be done with sign languages is an extension of the traditional notion of lemmatisation because phonetic and phonological variants, as well as morphological variants, must also be grouped

together under one lemma. Lemmatisation in this sense is necessary to conduct any linguistic research that relies on lexical contrast including studies of phonology, morphology and the lexicon.

As many sign language dictionaries function as translation dictionaries, few appear to follow principles based on this extended notion of lemmatisation.<sup>4</sup> One such attempt is the Auslan lexical database, available initially as a lemmatised print dictionary (Johnston 1989), later as an offline lexical database (Johnston 2001b), and most recently as Auslan SignBank, an openly publicly accessible online dictionary and as a restricted access lexical database (<http://www.auslan.org.au>). Other sign language dictionaries described as lemmatised resources include the Danish Sign Language dictionary (Centre for Tegnsprog 2008), the online New Zealand Sign Language dictionary (McKee et al. 2011), and a dictionary of German Sign Language (Deutsche Gebärdensprache, DGS) currently in progress (<http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/dictionary.html>) based on Konrad et al.'s (2012) lexical database. Sign language lexicographers working on these dictionaries have referred briefly to issues regarding lemmatisation (particularly regarding the treatment of polysemy and in distinguishing phonological and lexical variants) in their respective publications (e.g. Kristoffersen and Niemelä 2008; Troelsgard and Kristoffersen 2008; McKee and McKee 2013; Zwitterlood et al. 2013). However, the most comprehensive attempt to define lemmatisation practices in the context of sign languages can be found in Johnston and Schembri (1999) which is based on work conducted for the Auslan dictionary.<sup>5</sup>

Importantly, earlier dictionaries of BSL do not appear to have sufficiently addressed what may be called *the lemma dilemma* (cf. Brien and Turner, 1994). This is a major motivation for the development of a lexical database of BSL using the criteria set out in this paper, building on the work of Johnston and Schembri and others. For example, Johnston and Schembri (1999) note that, within the dictionary of BSL published by Brien (1992), modified variants of a single lemma are often listed separately without any indication of their relation to one another. Our own examination of Brien (1992) also revealed that signs which have more than one unrelated meaning (i.e. manual homonyms) are also listed as a single entry. Since signs are organised in this way within Brien (1992), it is very difficult to gain a clear understanding of the core lexicon of BSL from this dictionary alone. Johnston and Schembri (1999) note that if consistent lemmatisation principles were applied to this dictionary, then this would very likely have the overall effect of reducing the number of unique signs listed. In other words, not following lemmatisation practices presents a distorted view of the core lexicon.

It is important to note that the increase in sign language corpora which have only begun to be available in recent years have highlighted the absence of a lemmatised resource for many sign

languages. This has led sign language researchers to re-address the issue of lemmatisation and to develop such resources in parallel with corpus annotation. Prior to this, there was little attention given to the matter since many dictionaries were targeted at learners and thus were biased towards providing translation equivalents of words from a surrounding spoken language. Instead, sign language corpora require a system that identifies lexemes in a consistent way with a corresponding identifying gloss (i.e. an ID gloss, see below) so that more effective and reliable searches can be conducted.<sup>6</sup> In other words, one can search for and retrieve tokens of a given lexeme quickly since it has been consistently labelled throughout the corpus regardless of how it might have been modified. Without such a system, reports of lexical frequency of signs within a corpus (or any dataset) are inherently inaccurate and searching the corpus for all tokens of a given lexeme becomes a practically impossible task (for frequency lists based on lemmatised sign language corpora see Johnston 2012; Fenlon et al. 2014b). In the following section, we describe BSL SignBank and its association with the BSL Corpus and discuss the notion of an ID gloss in further detail before outlining our principles of lemmatisation.

### **3 British Sign Language and BSL SignBank**

#### *3.1 British Sign Language*

BSL is the language of the British deaf community. Whilst accurate statistics are difficult to obtain, a recent census report suggests that the number of signers in the United Kingdom is 15-20,000 (Office for National Statistics 2011). This figure is not without controversy and others have suggested that the number of signers is likely to be higher; the British Deaf Association report a figure of 120,000 extrapolated from a census report on the number of signers based in Scotland ([http://www.bda.org.uk/British\\_Sign\\_Language\\_\(BSL\)](http://www.bda.org.uk/British_Sign_Language_(BSL))). Based on our experience recruiting for the BSL Corpus, we are inclined to believe that higher numbers are unlikely. As with most sign languages in urban areas and industrialised countries, BSL's emergence is associated with the establishment in 1760 of the first deaf school in the UK. Since deaf children are very rarely born to deaf parents who sign, deaf schools play an important role in language transmission and act as a child's first point of exposure to a large community of signers. Although BSL has been the focus of linguistic research since the 1970s, there is still much about its structure and use that is poorly understood. As mentioned above, we currently do not have a dictionary that is representative of its lexicon. Recently, work has commenced on the creation of the BSL Corpus (Schembri et al. 2014) which contains a mix of naturalistic and semi-spontaneous data from 249 signers from around the UK. The BSL Corpus is intended to serve as a reference corpus enhancing our understanding of BSL at all levels of the language. However, the process of annotation at the lexical level is

dependent on the availability of a lemmatised resource to enable sign forms to be identified in a consistent way. As no such resource was available at the time of annotation, BSL SignBank was created to serve as a lemmatised resource including the core lexical signs in BSL; this is the focus of our paper here.

### 3.2 *BSL SignBank*

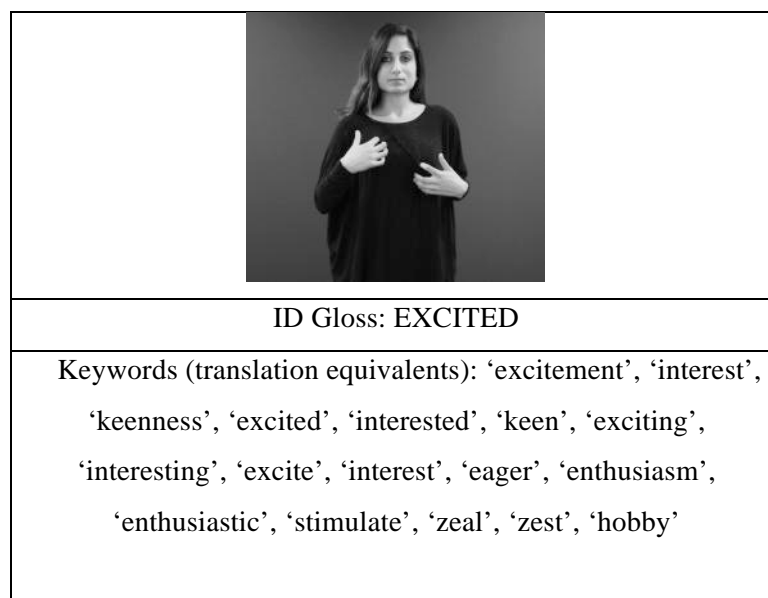
BSL SignBank (Fenlon et al. 2014a) is an electronic resource publicly available online that has a dual purpose as a lexical database for sign language researchers and as a dictionary for anyone with an interest in BSL.<sup>7</sup> In this paper, we focus primarily on its role as a lexical database (though see Section 5 for its role as a dictionary). Access to the lexical database (which provides additional information on each entry, see below) is available to researchers who register as university staff or as research students (<http://bslsignbank.ucl.ac.uk/accounts/register/>).

At the time of writing, BSL SignBank consists of 2,528 lexical signs of BSL (i.e. signs that are highly conventionalised in both form and meaning across the sign language community). Of these, roughly 1,700 sign types occur in (a) 50,000 sign tokens from the BSL Corpus conversation data, annotated as part of a lexical frequency study (Fenlon et al. 2014b) and a project on directional verbs (Cormier et al. 2014) and (b) 295 sign types for colours, numbers and countries from the BSL Corpus lexical elicitation data annotated as part of a study on lexical variation (Stamp 2013; Stamp et al. 2014). Additionally, (c) roughly 700 sign types in BSL SignBank are from Brien (1992) which did not occur in the corpus. Signs from Brien (1992) were added to BSL SignBank in order to ensure that the online dictionary would contain the core vocabulary of BSL, since entry into BSL SignBank prior to this point was contingent on a sign appearing in the portion of the BSL Corpus that had been annotated to date. BSL SignBank is a living resource in that it will continue to grow as further annotation work on the BSL Corpus is carried out (i.e. new lexemes that are encountered in the corpus are being added to BSL SignBank continuously).

For each lexeme within BSL SignBank, a set of keywords and a video showing the sign in its citation form is provided. For the purposes of this paper, we consider the citation form to be the lemma (i.e. the unmodified form of a given sign is used here as the headword of a lexeme). Decisions regarding which variants were considered citation forms are beyond the scope of this paper but are discussed in Cormier, Fenlon, et al. (2012a). The set of keywords are English translation equivalents for the meaning of the sign and are designed to act as a proxy for definitions and to cover its full range of meaning. Additionally available in the lexical database is the ID gloss and a phonological description of each entry. The ID gloss is a unique English-based translation used primarily as an annotation tag in the corpus for all occurrences of that lexeme regardless of

how it might be modified. It is important to note that the choice of the English word as an ID gloss for a particular lexeme is not meant to indicate the sign's core meaning or grammatical function. It is merely a label to uniquely identify each lexeme, to be used in annotation of sign language data, in lieu of any standardised orthography for the language. For the purposes of annotation, given the current state of technology with annotation software tools such as ELAN (<http://tla.mpi.nl/tools/tla-tools/elan/>, Wittenburg et al. 2006), it is useful to use ID glosses that have some meaningful connection to the lexeme, e.g. via one of the translation equivalents, since annotation is done by typing in the ID gloss.

In Figure 1 **Figure 1**, an example of a BSL lexeme with associated keywords, screenshot from its video and ID gloss is provided. One can see that the keywords associated with the lexeme ID glossed as EXCITED cover a range of meanings represented by different English translation equivalents and that the citation form associated with this sign is considered to be the two-handed variant (variants which are likely to be older, particularly those that can be assumed based on documented phonological processes, are often selected as citation form; see Cormier et al. 2012a).



**Figure 1.** The BSL lexeme EXCITED

In the following sections, we frequently refer to signs within BSL SignBank using the ID glosses assigned to them and in some cases also by the keywords associated with each lexeme. In each case, it is important to remember we are referring to the lexeme that the ID gloss represents. The lemmatisation principles that we describe in the next section are those that emerged in the course of developing BSL SignBank, building on Johnston and Schembri (1999), and are intended to serve as a guide to others undertaking similar tasks. The development of such principles requires regular discussions and consultations between sign language linguists and language consultants



from the local deaf community.

## 4 Principles of lemmatisation

In the following sections, we outline both straightforward and problematic cases that we encountered when lemmatising signs in BSL SignBank whilst outlining our basic criteria for lemmatisation. In Section 4.1, we describe criteria for distinguishing between phonological variants with similar or related meanings based on phonological form alone. In Section 4.2 and 0 on sign formation, we describe criteria for distinguishing between morphological variants. Section 4.4 focuses on using meaning/semantic criteria alone (e.g. determining when the meaning of a given token is different enough to warrant a separate lexeme) and, finally, in Section 4.5, we discuss lemmatisation with respect to non-manual features. Note that given the issues involved in attributing alterations in the form of a sign as inflectional or derivational (see Section 2.2), we use the alternative terms *sign modification* and *sign formation* respectively (cf. Johnston and Schembri 2007). Finally, although we discuss form and meaning criteria separately, it must be noted that it is always necessary to refer to both the sign's form and meaning to determine whether a given sign is a token of an existing lexeme within the database or a separate lexeme.

### 4.1 *Phonetic and phonological criteria*

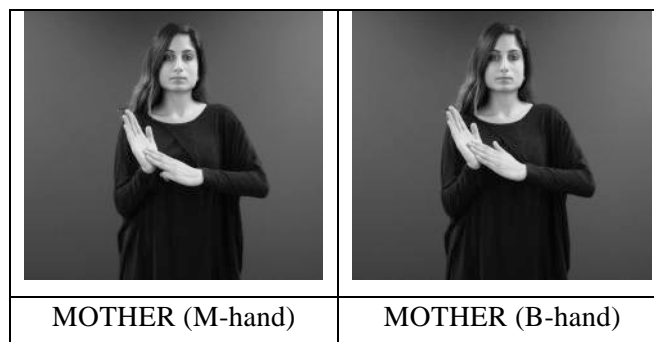
When determining the extent to which two variants differ from one another phonologically, it is necessary to refer to the parameters specified for each sign. Like words in spoken languages, signs can be analysed as composed of smaller units at the sublexical level. Sign language researchers frequently refer to handshape, location, movement, and orientation as the minimal contrastive parameters of sign production. Values within each parameter are often justified on the basis of minimal pairs. For example, the lexemes AFTERNOON and NAME differ in location whilst all other parameters are identical: AFTERNOON is produced with the same handshape, movement and orientation as NAME but is produced on the chin whilst NAME is produced on the forehead (see Figure 2). However, these signs can look different in connected signing or vary from signer to signer (e.g. NAME might be lowered slightly when signing quickly) just as words can vary in how they are pronounced (e.g. as mentioned in Section 2.2, the word *bath* has several pronunciations) (cf., Schembri et al., 2009). For each sign token, it is necessary to determine if its articulation significantly differs phonologically or phonetically from other tokens and to which lexeme it may belong.



**Figure 2.** Phonological contrast within the location parameter in BSL

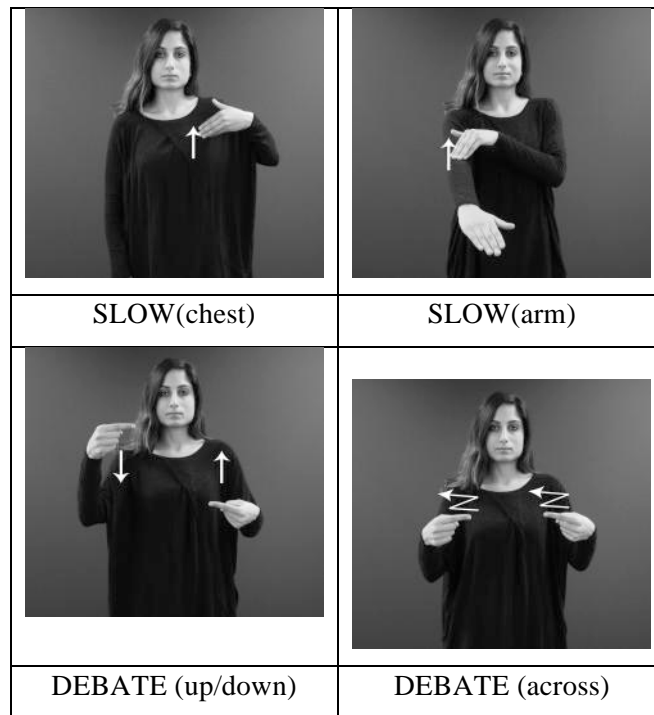
It is important to remember that it is rarely the case that decisions regarding separate lexemes are made with reference to form alone. On a basic level, decisions often must be made based on form and meaning. In this section, we outline our criteria using straightforward examples from BSL SignBank and then discuss various possibilities with similar/different forms and meanings.

*4.1.1. Phonological variants vs. lexical variants.* If sign tokens A and B differ in only one phonological parameter (e.g. handshape, location, movement, or orientation), and the meanings are the same or similar, then A and B are considered to be phonological variants. For example, the two variants of MOTHER shown in Figure 3 have the same meaning and differ only in handshape (i.e. the M-hand in which the index, middle and ring fingers are extended vs. the B-hand in which all fingers are extended).



**Figure 3.** Phonological variants of MOTHER (‘mother’, ‘mum’, ‘mummy’)

In Figure 3, these variants differ in the handshape parameter. Pairs of signs with related meanings can also differ according to the other parameters, e.g. location (SLOW) or movement (DEBATE) (see Figure 4).



**Figure 4.** Examples of phonological variants differing in location (SLOW ‘long’, ‘slow’, ‘slowly’) and movement (DEBATE ‘argument’, ‘debate’, ‘dispute’)

If sign tokens A and B differ in more than one phonological parameter, and the meanings are the same or similar, then A and B may be *lexical variants* (separate lexemes). For example, BSL NIGHT is produced with two B-hands in neutral space, and NIGHT2 is produced with a bent-V handshape at the nose, as shown Figure 5. (The ID glosses for lexical variants which have the same meanings are distinguished in BSL SignBank by a number suffix.)





**Figure 5.** Lexical variants NIGHT (‘evening’, ‘night’, ‘tonight’) and NIGHT2 (‘evening’, ‘night’, ‘tonight’)

4.1.2. *Phonological processes.* Straightforward instances of phonological variants that refer to a single lexeme include variants that are derived via documented phonological processes. Two phonological processes that are well documented are where some two-handed signs can be reduced to a one handed sign (known as weak drop), or that some one handed signs can become two-handed so that a symmetrical second hand is added (known as weak prop) (Brennan et al. 1984; Johnston 1989; Brentari 1998). Thus, signs which allow this process may be produced as one-handed or two-handed variants in connected signing without any modification of the sign’s overall meaning and the two variants may exist in free variation (some research indicates that the presence or absence of the non-dominant hand may be conditioned by the immediate phonological environment, e.g. Crasborn 2011). In BSL SignBank, one-handed and two-handed variants are always considered to constitute a single lexeme such as CAT and RIGHT in Figure 6.



**Figure 6.** One-handed and two-handed versions of the BSL lexemes CAT and RIGHT

Another phonological process relates to sign location. For example, the sign KNOW may be produced on the head, or lowered in connected signing. This lowering does not involve a change in meaning. Both are considered to be phonological variants of one lexeme, KNOW, as it is well known that signs can be lowered in connected signing and that this can be conditioned by a number of linguistic factors such as the location of the preceding or following sign, lexical frequency and grammatical category (e.g. Schembri et al. 2006; Schembri et al. 2009).

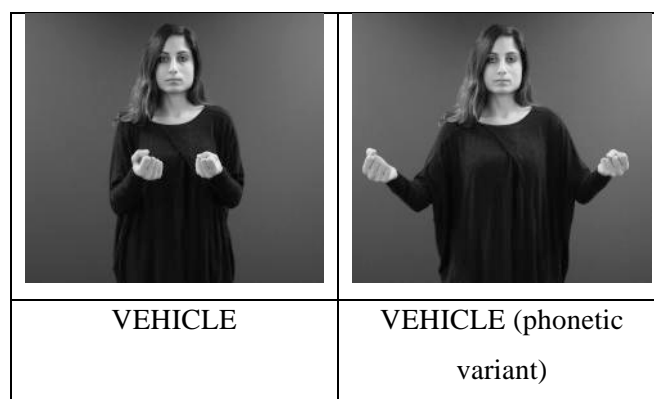
4.1.3. *Handshape variation/embellishment.* In some signs, the handshape can vary with no resultant change in meaning. One well-known example of such variation in BSL is the change in handshape in some signs to a 5 hand in the final phase of the sign. For example, SAME and KILL, that are both articulated with a 1-handshape throughout in citation form, can also occur in a form in which the handshape changes to a 5 handshape in the final phase (see Figure 7). We use the term *phonetic embellishment* to refer to this type of variation.<sup>8</sup>



**Figure 7.** Citation form and embellished forms for BSL SAME and KILL

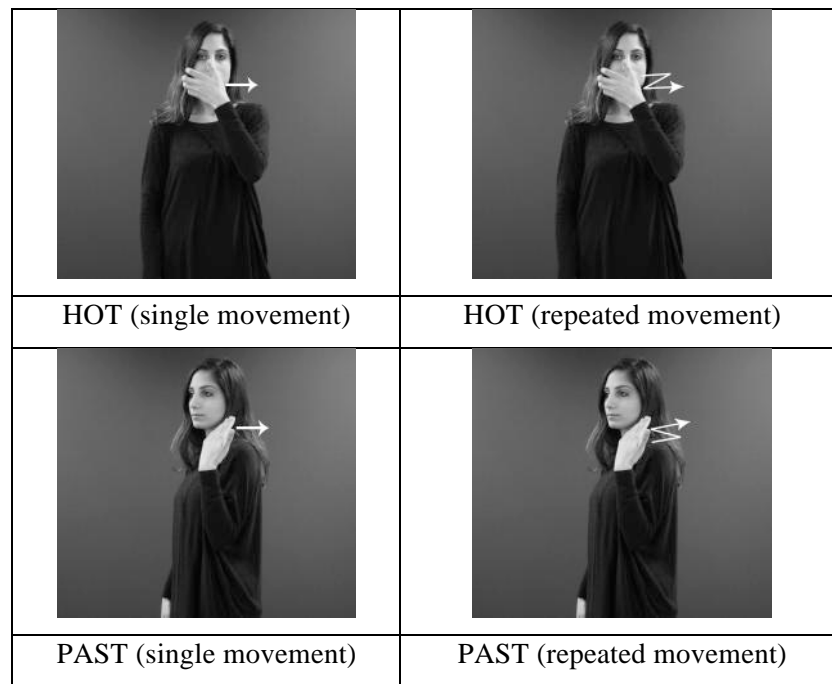
With other signs, formational variation may appear to result in a change in meaning but close attention must be paid to how consistent the form/meaning relationship is. For example, the BSL sign VEHICLE has the keywords ‘automobile’, ‘bus’, ‘car’, ‘driver’, ‘lorry’, ‘truck’, ‘van’ and

‘vehicle’ associated with it. The sign iconically represents holding the steering wheel of a vehicle with alternating movements up and down as if driving. Signers may produce the sign with hands further apart for ‘bus’ or ‘lorry’ (i.e. vehicles associated with larger steering wheels) as in Figure 8, but there is no clear and consistent difference between these two forms in BSL (e.g. signers may use a form with a larger distance between the hands for any kind of vehicle). Because the distance between the hands does not constitute a phonological parameter but is a phonetic difference that varies gradually, and there is no evidence of consistent form/meaning differences, these are considered variants of one lexeme, VEHICLE.



**Figure 8.** Example of phonetic gradience in the articulation of the BSL lexeme VEHICLE

*4.1.4. Repetition.* Particular care must be taken when considering signs which may vary in terms of the number of movement repetitions. If the meanings of these variants are the same or similar, then they are likely to be phonetic variants. Examples of signs which may be repeated are HOT and PAST (see Figure 9). In some cases, the repeated variant might be expected to co-occur more with mouthings corresponding to keywords that are multisyllabic (e.g. ‘heating’ and ‘recent’), but it is not clear to what extent this difference is consistent and there are likely counterexamples. Therefore, pairs of signs that vary according to repetition alone are considered variants of one lexeme.



**Figure 9.** Examples of the BSL lexemes HOT and PAST without and with repetition

In this section, phonological criteria for lemmatisation have been outlined. In the following sections, additional criteria for lemmatising signs according to morphological processes are outlined beginning with sign modification.

#### 4.2 *Morphology: sign modification*

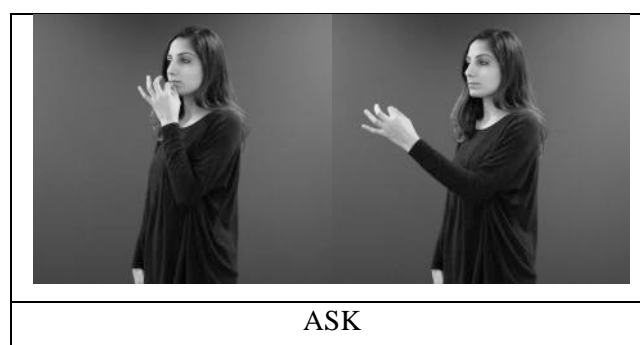
Our practice of lemmatisation involves grouping together morphological variants that are the result of sign modification – in other words, modifications of existing/known conventional signs that augment/change the meaning of the stem in predictable ways, both formationally and semantically, in sets of signs. In the case of BSL, variants marked for number, aspect, and what has come to be known as agreement in the literature (Padden 1988) are all assigned to the same lemma (representing the unmodified form). Examples of each are provided below.

*4.2.1. Aspect marking.* Aspect marking is typically used with verbs or other predicates to represent how an action, event or state unfolds over time. For example, two variants of the lexeme WALK mean ‘to walk’ and ‘to walk for a long time’ respectively (the latter having a larger circular motion associated with durational aspect). In some cases, a token may appear to be an aspectual modification of a lexeme but instead has a specific meaning that cannot be predicted on the basis of the modification alone. For example, DRINK, with a single movement toward the lips, has the meaning ‘to drink any type of beverage’ whether it is a glass of water, juice, or alcohol. Durational aspect marking, e.g. with a larger, repeated, inward motion, can be added to this sign to indicate

drinking over a period of time. A related variant with slightly different movement (i.e. smaller, outward circular motion along the horizontal plane that also indicates that the action was conducted over a period of time) means ‘to drink alcohol excessively’ or ‘one who drinks alcohol excessively’ rather than ‘to drink any type of beverage excessively’. It is this specific meaning, which cannot be predicted from the addition of durative aspect to DRINK alone, that warrants a separate lexeme status for this variant, ALCOHOLIC. Lexemes like ALCOHOLIC demonstrate that not all tokens should be grouped together based on modifications to underlying forms alone but careful attention should be paid to any difference in meaning. If the change in meaning is unpredictable (as is the case in ALCOHOLIC), then separate lexeme status is justified.

*4.2.2. Directional modification.* Some verbs can be spatially modified so that they are directed towards the physically present referents of their arguments, or locations associated with absent referents. These verbs are known as agreement verbs (e.g. Padden 1988), directional verbs (e.g. Baker-Shenk and Cokely 1978), or indicating verbs (e.g. Liddell 2000). There is no justification in assigning separate lexeme status to each directional variant as the resulting change in meaning is predictable (i.e. whatever its direction, the overall verbal meaning of the sign remains the same). Thus, these variants are all considered to constitute a single lexeme.

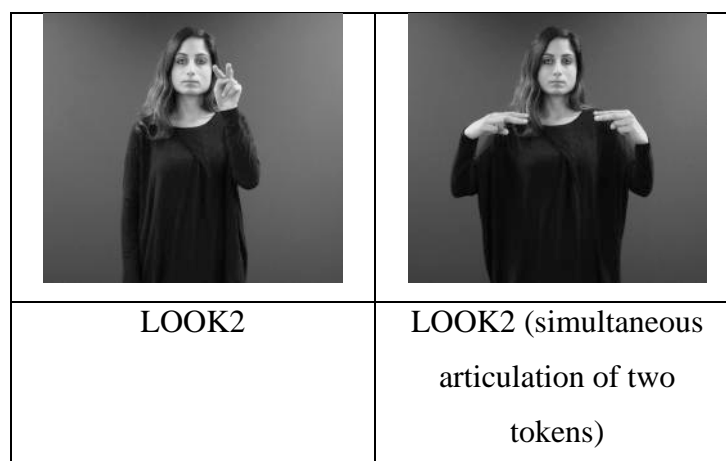
*4.2.3. Number and distributive marking.* Sign languages have a number of options available to mark number. For some nouns, this might involve a simple repetition of the sign along a short sideways path movement (primarily for signs produced in neutral space). Verbs (such as the verb ASK, shown in Figure 10 in citation form) can be modified for number/distribution using two different types of marking: the multiple and the exhaustive. Both types differ in movement and meaning: the former involves multiple repetitions of the verb towards several locations in space with a distributive reading (e.g. ‘ask each’); the latter involves a single sweeping movement across the signing space at the end of the sign and has a general plural or exhaustive reading (e.g. ‘ask more than one’ or ‘ask all’). Since the difference between these variants is always predictable, we do not consider these variants as separate lexemes.





**Figure 10.** BSL ASK in citation form

Another way that signers may mark multiple referents is by using both hands instead of one (Johnston and Schembri 1999). This differs from the process mentioned earlier in which some one-handed signs can be articulated with two hands to produce a symmetrical sign (as with CAT in Figure 6). Here, some one-handed signs (usually verbs) can be produced using both hands to indicate more than one referent (and, in some cases, to give a reciprocal meaning to the verb) rather than being a purely phonological element. For example, the sign LOOK2 is entered into BSL SignBank as a one handed sign. In cases where it is produced with both hands in connected signing to mean ‘two people looked’, both hands are considered independent signs since both represent different entities. Alternatively, the hands may be oriented towards each other to give the meaning ‘two people looked at each other’ as in Figure 11. In each case, the resulting two-handed articulation is not treated as separate lexeme as there is no significant change in meaning.

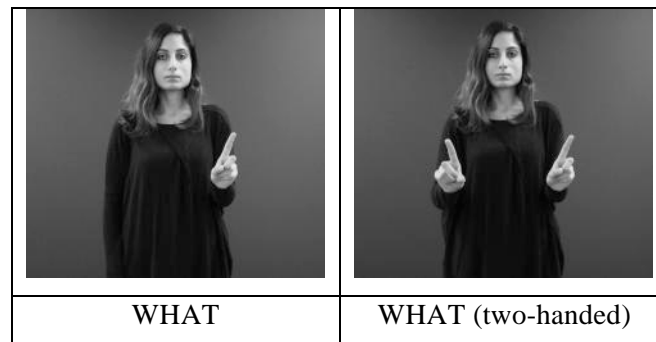


**Figure 11.** One-handed LOOK2 vs. simultaneous articulation of two tokens of LOOK2 (the latter expresses the meaning ‘two people looked at each other’)

*4.2.4. Intensification.* The movement in some signs can be modified so as to intensify the meaning in some way. For instance, the signs RED and QUICK can be modified so that the beginning of the sign is held for longer than is usual to mean something is ‘very red’ or is ‘very quick’. Generally speaking, this type of modification is not considered a separate lexeme. In other words, forms corresponding to ‘red’ and ‘very red’ will come under the same lemma RED as the two variants are formationally similar (i.e. they differ in only one parameter) and the change in

meaning is predictable.

In some cases, a sign's meaning can be intensified (e.g. for emphasis) by using both hands as in Figure 12. As mentioned previously, two-handed variants generally do not warrant a separate entry into a lexical database; and the use of both hands as a marker of intensification is treated no differently here.



**Figure 12.** WHAT vs. WHAT with two hands

### 4.3 Morphology: sign formation

In this section, we discuss criteria for distinguishing lexemes on the basis of sign formation. By sign formation, we refer to processes by which a sign form appears to have been derived from another sign and the resulting form expresses a new meaning which may not be fully predictable (as it would be with sign modification). In addition, this sign form could itself serve as a ‘stem’ which could be modified. In these cases, it is often justified to recognise two different lexemes on this basis; the difference between these related forms may correspond roughly to what some consider to be derivational morphology in the literature.

4.3.1. *Noun vs. verb alternation.* One well-known example of sign formation commonly referred to in the sign language literature is the movement alternation observed in related noun and verb pairs in some sign languages (Supalla and Newport 1978; Sutton-Spence and Woll 1999). In these pairs, the movement associated with the noun is shorter and repeated whilst the movement associated with the verb is longer and not repeated, as with signs like ASL FLY and AEROPLANE as in Figure 13, based on Supalla and Newport (1978).

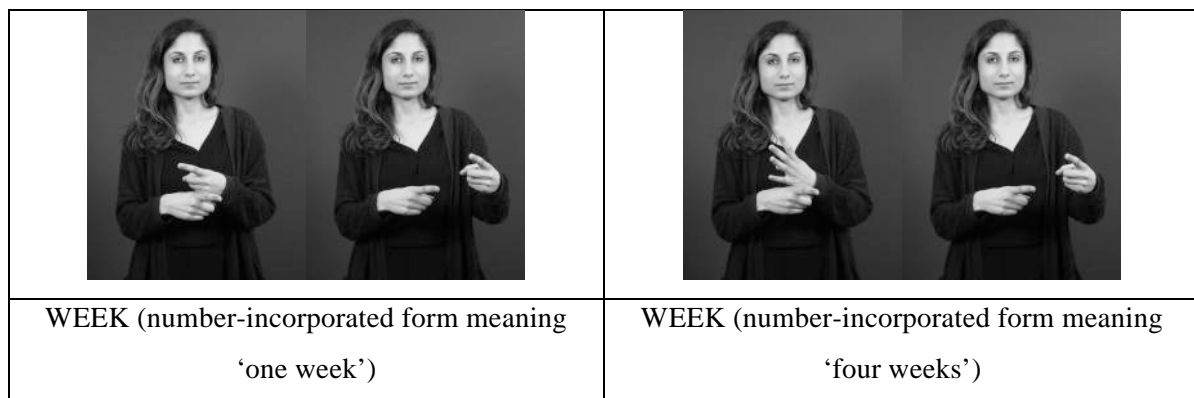


**Figure 13.** ASL FLY and ASL AEROPLANE

Despite claims by Sutton-Spence and Woll (1999), the extent to which this distinction exists in BSL is unclear (cf. Johnston, 2001, for Auslan). Some related noun and verb pairs appear to differ systematically based on movement alone. These include pairs like KEY and LOCK, and DOOR and OPEN-DOOR. Many forms do not appear to have a systematic difference in movement when functioning as nouns or verbs, e.g. AEROPLANE, TEACH and FINGERSPELLING. As a rule, only related pairs where the movement alternation appears to be systematic are treated as separate lemmas. This appears to be the case for KEY/LOCK where LOCK involves a single forearm rotation movement as in a locking action and KEY involves a repeated forearm rotation. In BSL, as in Auslan, it is often difficult to determine whether this alternation appears to be systematic

for a given pair as there is not a clear consensus amongst signers as to whether one exists for all possible noun-verb pairs (Schembri et al. 2000; Johnston 2001a). Further analysis using the corpus data together with elicited data will determine this and current decisions may need to be revisited.

*4.3.2 Number incorporation.* Some signs in BSL can incorporate number signs. In some cases, the underlying sign is considered a bound morpheme which must be combined with a number sign in order to be a fully articulated sign. One example of such a sign is WEEK, as shown in Figure 14. Here, the number is represented by the handshape (i.e. in this case, the number ONE is indicated by the extended index finger) while the remaining elements constitute a bound morpheme that cannot exist on its own. In each case, all related tokens are represented by the same lemma. In other words, WEEK always represents the resulting combination of the two morphemes.

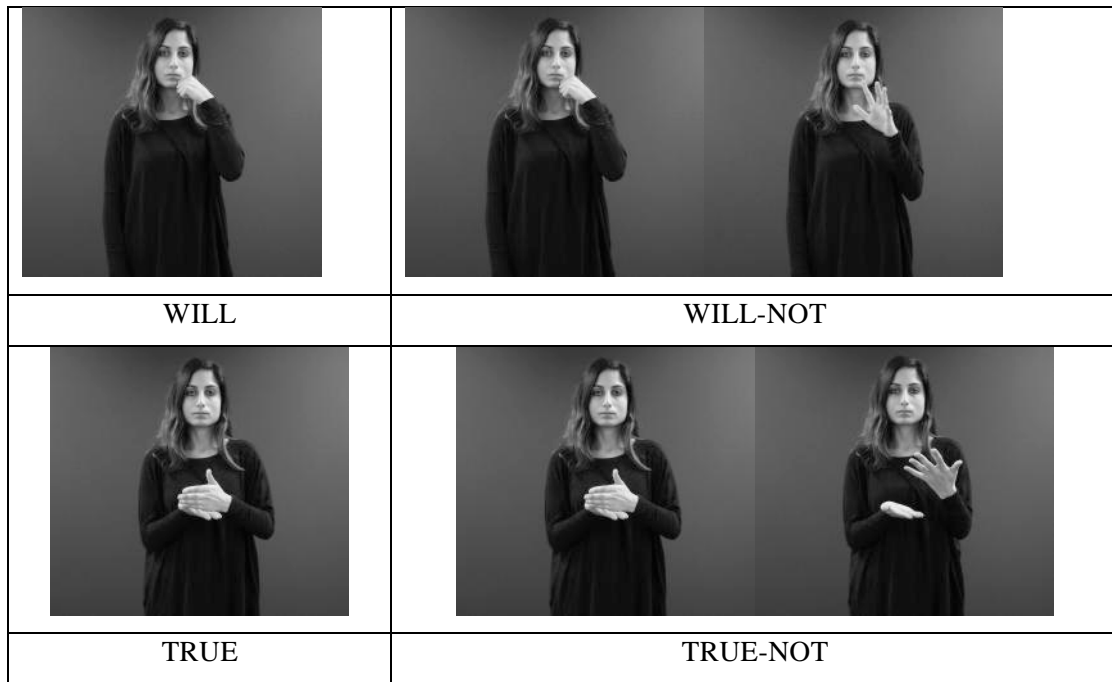


**Figure 14.** Examples of number incorporated forms of WEEK

In all cases, we do not recognise each possible form (e.g. WEEK-TWO, WEEK-FOUR) as different lexemes partly because the change in meaning is predictable. Also, in some cases, the number of possible combinations would be numerous and it is simply not economical to enter each individual form combination into the database (this would rapidly exaggerate the number of unique entries listed). Instead, single entries (e.g. WEEK in Figure 14) are added to the database along with information that this form could incorporate number.

*4.3.3. Negative forms.* Another type of sign formation that occurs in sign languages is negative modification. In BSL, the negative marker is an open 5 hand with forearm rotation movement. A small set of signs in BSL can be modified for negation; this set includes the lexemes WILL, WANT, AGREE, and TRUE. For lemmatisation, negative variants are recognised as separate from the underlying base variant, are treated as unique lexemes, and are entered separately into the database. The examples listed above are all listed as WILL-NOT, WANT-NOT, AGREE-NOT, and TRUE-NOT in BSL SignBank; see Figure 15 for examples. The approach taken here is

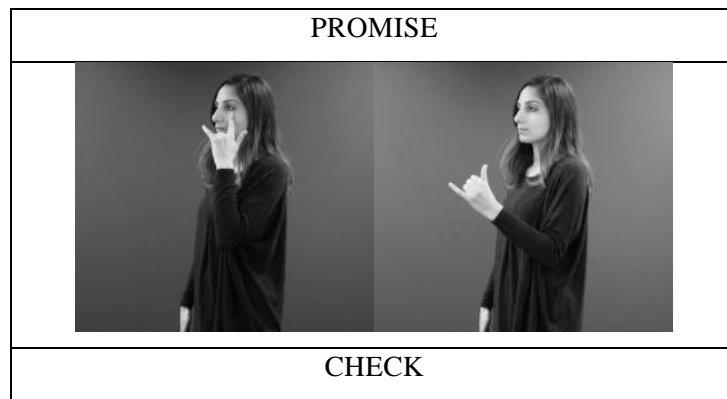
in contrast to the approach taken for number incorporation since the number of possible combinations involving negative forms appears to be much smaller than the number of possible combinations involving number incorporation (even though, in both cases, the change in meaning appears predictable).



**Figure 15.** WILL ('shall', 'will', 'would'), WILL-NOT ('refuse', 'won't', 'wouldn't'), TRUE ('fact', 'true', 'truth'), TRUE-NOT ('false', 'falsehood', 'untrue')

*4.3.4. Compounds and collocations.* Compounding refers to the process of creating new signs from two or more free morphemes and is a process attested in many sign languages documented to date (Klima and Bellugi 1979; Johnston and Schembri 1999; Sutton-Spence and Woll 1999). In each case, compounds are considered unique lexemes and are entered into the database as such. Examples of compounds in BSL include PROMISE (SAY+TRUE), BLOOD (RED+SPREAD) and CHECK (SEE+MAYBE) as in Figure 16.



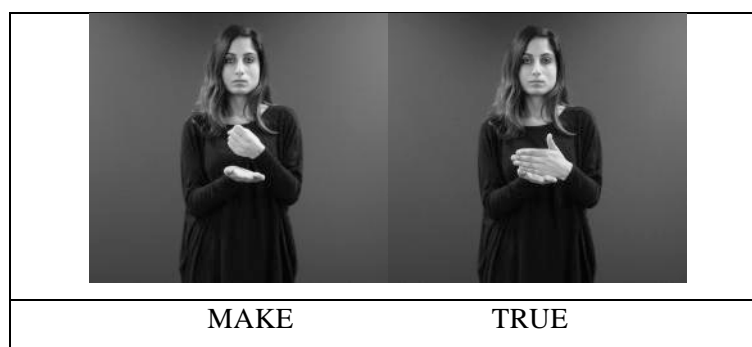


**Figure 16.** PROMISE (‘honest’, ‘promise’, ‘truth’) and CHECK (‘check’, ‘examine’, ‘test’)

One challenge for the sign language lexicographer is to determine when two signs represent a compound (and therefore a unique lexeme in SignBank) as opposed to when they represent a collocation (i.e. two signs that appear next to one another frequently but do not represent a lexeme). Collocations refer to words that frequently occur together such as ‘strong tea’ or ‘crystal clear’ in English. Similar examples from BSL include the sign pair MAKE and TRUE (see Figure 17) which also represent concepts that frequently collocate in English (i.e. ‘make sure’: PT:PRO1SG WANT MAKE TRUE, ‘I want to make sure’). In connecting signing, pairs of signs that are frequent collocates can often appear like compounds in that they are produced as if they are a single monosyllabic sign. For example, MAKE followed immediately by TRUE is often articulated as a single monosyllabic sign with the handshape and orientation of the non-dominant hand in the sign TRUE anticipated at the start of the sign (compare MAKE and TRUE in Figure 18 with PROMISE in Figure 16).



**Figure 17.** Citation forms for MAKE (‘create’, ‘make’, ‘manufacture’) and TRUE (‘fact’, ‘true’, ‘truth’)



**Figure 18.** MAKE and TRUE produced with a high degree of assimilation

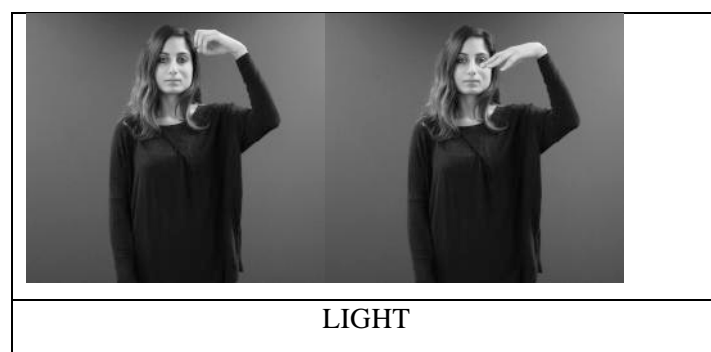
Although researchers have noted that the compounding process may alter the articulation of the sign so that it becomes a monosyllabic sign (Brentari 1998), it is clear from collocates like MAKE and TRUE that such phonological processes are not limited to compounds alone. Alternative criteria for distinguishing between compounds and collocates may be based on meaning. In other words, the meaning derived from the combination of two morphemes within compounds is often not predictable. BSL CHECK (a compound of SEE and MAYBE) has a range of English translation equivalents not predicted by the combination of SEE and MAYBE (see Figure 16). In some cases, the resulting compound may have taken on a broader or more specific meaning. For example, BLOOD is a compound of RED and SPREAD. Separately these two signs could refer to anything red that is spreading (e.g. spilled ketchup, red lava, etc.) but as a compound this sign refers only to blood. The fact that the resulting compound has an unpredictable meaning is an important point in support of the fact that these signs constitute separate lexemes (this is also standard practice in lemmatised lists representing words from English, cf. Sterkenburg 2003). In comparison, whilst collocational pairs may look like compounds, their meaning remains predictable. That is, the meaning that is derived from the combination of the signs MAKE and TRUE is not novel and can easily be predicted. Only when such combinations take on broader, unpredictable meanings can we consider them as unique lexemes.

*4.3.5. Signs relocated in space.* Some signs can be relocated easily to different locations on the body or in the signing space sometimes with a slight change in meaning and sometimes forming a unique lexeme in its own right. For example, EMIT can be located and oriented in different ways for various meanings linked to emitting something (e.g. light or sound) as in Figure 19.



**Figure 19.** EMIT (‘emit’, ‘illuminate’, ‘radiate’) from different locations

The examples in Figure 19 are considered to represent one lexeme, the general sign EMIT. This is because in each case, this token has not acquired a specific meaning. In other words, it continues to mean ‘something (a light source/sound source/etc.) emitted from somewhere’. However, if the sign EMIT produced in high neutral space with the palm facing downwards (as in Figure 20) is presented to a BSL signer, it is highly likely that the signer will indicate that this sign has a specific meaning of ‘an overhead light’. Although this sign could feasibly refer to anything being emitted from that location (e.g. a speaker facing downwards), it has acquired a specific meaning and therefore is considered a separate lexeme.



**Figure 20.** LIGHT (‘illuminate’, ‘lamp’, ‘light’)

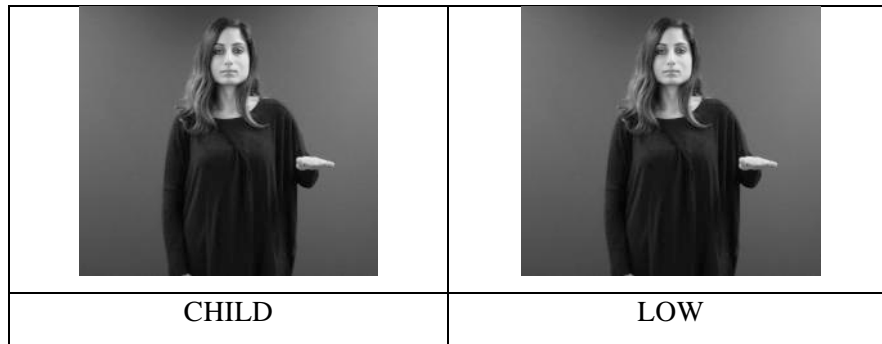


Likewise, OPERATE can also be used productively in the same way as EMIT to refer to operating on any part of the body (e.g. by moving it to any body part such as the arm, neck, chest). However, when it is produced in certain locations, the overall form and meaning pairings appear more conventionalised. For example, when the sign is produced on the signer's right side of the waist, it has the specific meaning of 'appendectomy'. As a result, it is identified as a separate lexeme: APPENDECTOMY.



**Figure 21.** OPERATE ('operate', 'operation', 'surgery') and APPENDECTOMY ('appendectomy', 'operate', 'operation', 'surgery')

4.3.6. *Lemmatisation based on different sign modification patterns.* If we have two variants with related meanings that are similar in form, the argument for recognising two separate lexemes may be based on whether one variant can take different morphological modifications. For example, both CHILD and LOW are formationally similar in citation form and have a similar general meaning of (something of) a low-level height. However, these signs can be modified differently. CHILD can be modified to indicate number (with a movement to the side) whilst LOW cannot (see Figure 22). Additionally, LOW can be modified for intensification with a longer downward movement (i.e. to mean something is very low); this kind of modification is not observed for CHILD. The fact that both take different types of modification is indicative of the fact that they represent different parts of speech: CHILD typically functions as a noun and LOW typically functions as an adjective. Similar claims can be made for the pair ADULT/HIGH.



**Figure 22.** CHILD (‘child’, ‘kid’, ‘young’) and LOW (‘little’, ‘low’, ‘short’)

Similarly, WOOD3 and HARD are identical in form (repeated movement of the thumb against the palm of the other hand; see Figure 23) and have related meanings (the quality of hardness). However, WOOD3 can be used to refer to any wood regardless of degree of hardness. This distinction in meaning justifies recognising these two forms as separate lexemes. Additionally, they are also considered to be separate lexemes because HARD can be modified in a way to mean ‘very hard’ via a change in the movement (i.e. single movement of both hands downward) and this does not apply to WOOD3.



**Figure 23.** WOOD3 (‘wood’, ‘wooden’) and HARD (‘difficult’, ‘hard’, ‘problem’)

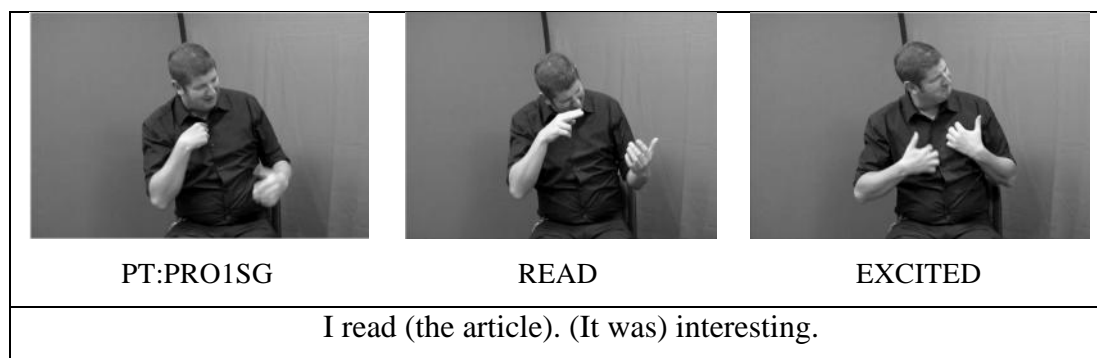
#### 4.4 Meaning

In addition to the form of the sign, as noted previously, it is also necessary to refer to the sign’s meaning when determining whether two variants constitute the same lexeme or not. This is clearly required for homonyms: pairs of sign that have the same phonological form but differ in meaning. For example, both BSL BROTHER and MARCH-MONTH are produced with two fist hands in neutral space brushing against each other with alternating up-and-down movement (Figure 24).



**Figure 24.** Homonyms – BROTHER (‘brother’) and MARCH-MONTH (‘March’)

As the meanings in each case are distinct, BROTHER and MARCH-MONTH are treated as homonyms and therefore separate lexemes. In some cases, the meaning represented by a single lexeme can also appear very broad as opposed to being clearly distinct as with BROTHER vs. MARCH-MONTH. For example, EXCITED is polysemous and displays a very broad range of meanings, including ‘ambitious’, ‘eager’, ‘eagerness’, ‘excited’, ‘hobby’, ‘interested’ and ‘motivated’. These meanings are all related and they all share the same sign form (i.e. there is no difference in form when this sign is used to mean ‘excited’ or ‘interested’) as indicated in Figure 25 and Figure 26.



**Figure 25.** Use of EXCITED to mean ‘interesting’





**Figure 26.** Use of EXCITED to mean ‘excited’

There are many cases in BSL SignBank where a single lexeme has several associated keywords demonstrating a broad range in meaning. TAKE includes the keywords ‘adopt’, ‘adoption’, ‘burglary’, ‘burglar’, ‘get’, ‘grab’, ‘nick’, ‘obtain’, ‘pinch’, ‘rob’, ‘robber’, ‘robbery’, ‘seize’, ‘shoplift’, ‘snatch’, ‘steal’, ‘stealing’, ‘take’, ‘theft’, ‘thieve’. Similarly BALL includes the keywords: ‘ball’, ‘football’, ‘globe’, ‘global’, ‘round’, ‘sphere’, ‘spherical’, ‘world’. In each case, there is no difference in the citation form of the sign used for each meaning and there is an obvious shared meaning between these keywords – e.g. the keywords all refer to something ball-shaped. Since these meanings are all related and they all share the same sign form, they are considered in each case to be part of one lexeme.

The meaning range of a polysemous sign can extend to both abstract and concrete senses; often this is due to metaphorical extension of a concrete sense to an abstract sense. For example, the keywords associated with the lexeme ROAD indicate both concrete senses (‘avenue’, ‘path’, ‘road’, ‘street’) and abstract senses (‘method’, ‘way’). Likewise, DESIRE is produced at the neck and can mean ‘thirsty’. As in the English word ‘thirst’, the use of DESIRE is not only applied to being thirsty for liquid nourishment but for other things as well, such as knowledge. Likewise, in BSL, it is also used to indicate desire of any kind. To represent this extension in meaning, DESIRE has a number of keywords such as ‘desirable’, ‘desire’, ‘dry’, ‘fancy’, ‘thirst’, ‘thirsty’ and ‘wish’. Again, as there is no difference in the form of the sign according to the senses displayed by the keywords, and because ‘desire’ is a metaphorical extension of the meaning ‘to be thirsty’, all of these keywords are assigned to one lexeme, DESIRE.

In many cases, the meanings exhibited by the keywords listed in BSL SignBank indicate words that have a superordinate/subordinate relationship to one another. A clear example of this can be seen with the sign BIRD which includes the keywords ‘bird’, ‘birdie’, ‘chicken’, ‘fowl’, ‘pigeon’. This list includes keywords that are co-hyponyms (e.g. ‘chicken’ and ‘pigeon’) as well as




keywords that are superordinate/hypernyms such as ‘bird’. As there is no formational difference in the sign when used to refer to either a chicken or a pigeon or to any bird, there is no justification in acknowledging these items as separate lexemes.





In other cases, meanings of keywords in BSL SignBank have a metonymic relationship to each other. Examples include BADGE with the keywords ‘badge’, ‘delegate’, ‘official’, ‘representative’, ‘steward’, and ‘supervisor’, as well as MEXICO with ‘Mexican’, ‘Mexico’ and ‘sombrero’. Our guiding principle is, if two forms are identical and the meanings are related, then they refer to the same lexeme. Generally, this means that keywords which have a metonymic relationship with one another are grouped together under the same lemma. However, this is not always the case. For example, the sign COMMUNIST has keywords ‘communism’, ‘communist’, ‘feminism’, ‘feminist’, ‘militant’, ‘radical’, ‘socialism’, ‘socialist’, but is a separate lexeme from the identically formed sign RUSSIA which has the keywords ‘Russia’, ‘Russian’, ‘Soviet’, ‘Soviet Union’, ‘USSR’. One reasoning for this is because RUSSIA is a sign name which is highly institutionalised and therefore is recognised as a separate lexeme. Following this line of reasoning, one may argue that the keywords associated with MEXICO are indicative of two separate lexemes: MEXICO and SOMBRERO. However, this is not the case because there is little evidence that the sign which may be glossed as SOMBRERO is a conventionalised sign (i.e. is a conventionalised form/meaning pair) in BSL. For example, if a BSL signer were presented with this sign (without mouthing) and asked to tell us what it meant, it is unlikely that their first response would be ‘sombrero’ but most likely would be ‘Mexico’. A similar argument can be made for CHRISTMAS and BEARD2. Since the sign CHRISTMAS is an iconic depiction of Father Christmas’ beard, one could argue that the separate lexeme status accorded to both CHRISTMAS and BEARD2 is not justified because the extension in meaning is similar to that amongst the keywords in MEXICO. However, although the sign CHRISTMAS is likely to have started as an extension of the sign BEARD to refer to Father Christmas as a sign name, it has now taken on a more general meaning (i.e. refers to Christmas, the holiday) and is highly conventionalised within the signing community (i.e. we would expect that signers would be just as likely to respond ‘Christmas’ and ‘beard’ when asked what this form means).<sup>9</sup> There are many, many other examples like these in BSL SignBank and, in many cases, it can be difficult to decide when meanings are different enough to constitute separate lexemes (i.e. homonyms). In all cases, both form and meaning must be considered (see also Section 4.4.1 below).

It has been observed that the lexicon of a sign language may expand through a process known as loan translation (Battison 1978). Loan translations appear to be motivated by related meanings existing within BSL homonyms, where one or more of the signs involve a loan translation from

English. For example, both the BSL signs HANGING and HANGOVER depict someone being hanged by the neck. With HANGOVER, this is not an iconic depiction, because the form has no bearing on the meaning of being ill due to excessive drinking. Instead, what has happened here is that semantic extension has occurred with HANGING based on the English word ‘hang’ which comprises part of the word ‘hangover’. It is standard practice within BSL SignBank to recognise these entries (i.e. HANGING and HANGOVER) as separate lexemes provided they are considered to be in widespread use within the community (i.e. they are not restricted to one individual and occur several times across individuals within the corpus).

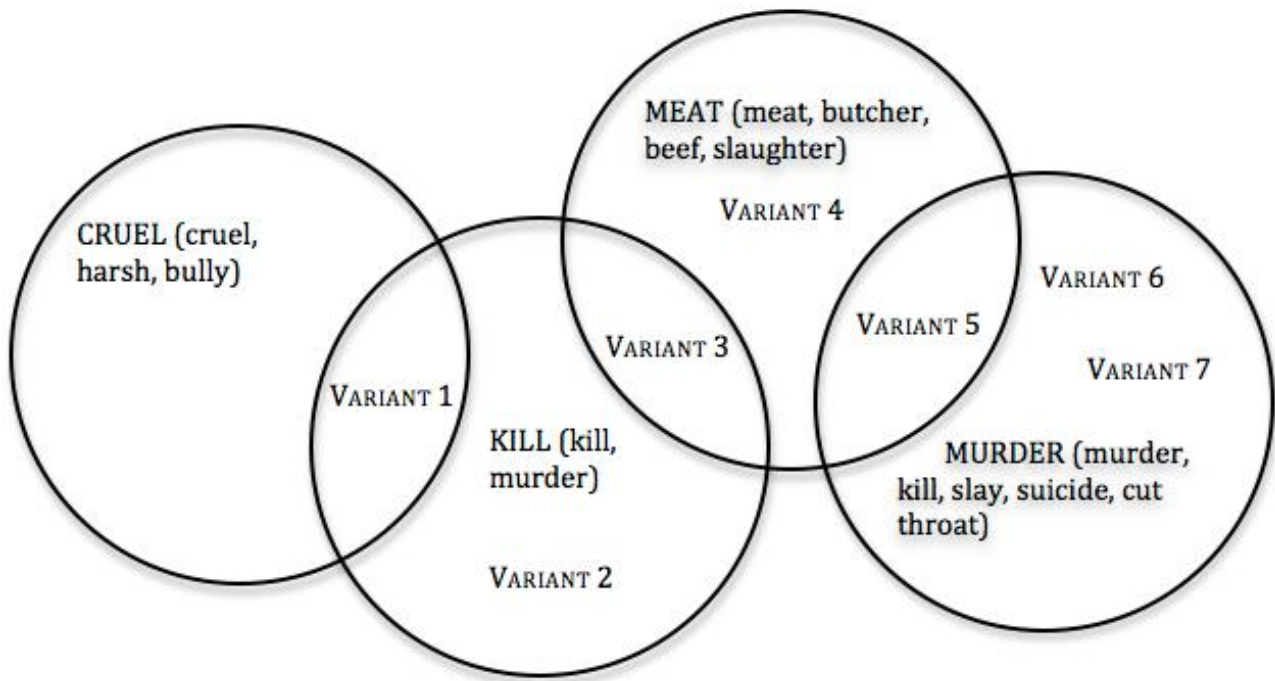
4.4.1. *Multiply related variants: the case of CRUEL/KILL/MEAT/MURDER.* Relationships between form and meaning across different variants can be quite complex. Here we describe one of the most complex set of relationships that we encountered in the course of annotating the BSL Corpus. There are seven BSL sign variants involving meanings linked to ‘cruelty’, ‘killing’, and ‘meat’, all produced at the neck, with overlapping/related forms and meanings (Figure 27).

		
<p style="text-align: center;">VARIANT 1</p> <p>1-hand with extended index finger at neck with forearm twist (‘bully’, ‘cruel’, ‘harsh’, ‘kill’, ‘murder’)</p>	<p style="text-align: center;">VARIANT 2</p> <p>1-hand with extended index finger touching neck opening to 5-hand away from body (‘kill’, ‘murder’)</p>	<p style="text-align: center;">VARIANT 3</p> <p>1-hand with extended index finger touching neck (‘butcher’, ‘beef’, ‘kill’ ‘meat’, ‘murder’, ‘slaughter’)</p>

			
<p>VARIANT 4</p> <p>1-hand with extended index finger touching neck repeatedly ('butcher', 'beef', 'kill' 'meat', 'murder', 'slaughter')</p>	<p>VARIANT 5</p> <p>Fist hand with extended thumb touching neck repeatedly ('butcher', 'beef', 'cut throat', 'kill' 'meat', 'murder', 'slaughter', 'slay', 'suicide')</p>	<p>VARIANT 6</p> <p>Fist hand with extended thumb moving across throat ('cut throat', 'kill', 'murder', 'slay', 'suicide')</p>	<p>VARIANT 7</p> <p>1-hand with extended index finger moving across throat ('cut throat', 'kill', 'murder', 'slay', 'suicide')</p>

**Figure 27.** Seven sign forms linked semantically and phonologically in overlapping ways

The way we have lemmatised these variants is shown in Figure 28 by grouping together similar forms and meanings. These seven variants form four separate lexemes (CRUEL, KILL, MEAT and MURDER), each with related phonetic/phonological variants. At the endpoints of the chain, CRUEL (including only Variant 1) and MURDER (including Variants 5, 6 and 7) look quite different from each other and have different meanings. But by considering related lexemes KILL and MEAT (and their phonetic/phonological variants, i.e. Variants 1 through 5), we can see how they are related to each other, both in form and meaning. These multiply related variants highlight the importance of considering all possible related variants (in both form and meaning) in the process of lemmatisation.



**Figure 28.** A chain of phonologically and semantically related variants

#### 4.5 *Mouthing and other non manual features*

In this section, we describe the influence of mouthing and non-manual features on decisions regarding lemmatisation. In connected signing, signs in BSL are often accompanied by the silent mouthing of English words. Work on other signs languages, such as the related variety Auslan (Johnston, van Roekel and Schembri, in press), show that this mouthing is a spoken/signed language contact feature which exhibits considerable individual variation, but it tends to follow specific patterns (signs used as nouns, for example, are much more likely to be accompanied by mouthing than verbs). In our work on BSL, whether a variant constitutes a lexeme is decided independently of non-manual features which may accompany a sign. For example, the English mouthing ‘forest’ frequently accompanies the sign TREE when modified for number (i.e. it has a short sideways movement repeating the sign). Lemmatisation of these forms follows that of number and distributive marking as noted above in Section 4.1.3, regardless of mouthing. Strictly speaking, mouthing is only used as a cue for possible meanings in context (e.g. to develop the list of associated keywords or definitions) rather than as justification for lemmatisation. This is true even if one mouthing is more likely to be used with a particular variant than another, as in the case of two variants of the lexeme WORTH (‘appreciate’, ‘value’, ‘valuable’, ‘valued’, ‘worth’, ‘worthwhile’) where one variant is initialised with the letter -W- and may be more likely to occur with mouthing ‘worth’ or ‘worthwhile’ than the other keywords.<sup>10</sup>

Frequently, some signs can be modified for meaning using non-manuals features such as the



mouth or facial expressions. The types and number of mouth gestures that are attested in BSL are not well understood although Sutton-Spence and Woll (1999) describe eight manner and degree mouth gestures that are important in BSL. For example, the verb DRIVE can be articulated with a specific adverbial mouth gesture to produce a sign meaning ‘to drive carelessly’ (see Figure 29) (Liddell 1980; Lewin and Schembri 2011).

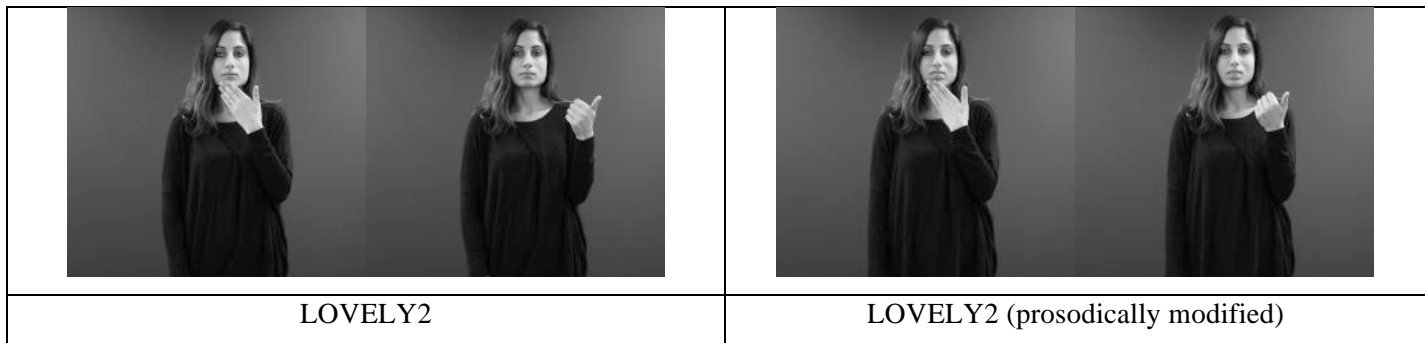


**Figure 29.** DRIVE (‘drive’, ‘driving’) and DRIVE with th-mouth gesture (meaning ‘to drive carelessly’)

In cases such as these, it is not possible to recognise separate lexemes with every verb that has been modified with a mouth gesture – therefore, the variants shown in Figure 29 are represented by the lemma DRIVE. This and other mouth gestures which have a morphological function are thus treated the same as other sign modifications described in 4.2 above. The same practice is applied to variants of signs that are formationally similar but differ in whether they express a negative or positive meaning based on facial expression/mouthing alone. For example, the sign EMOTION is understood to represent the full range of emotions. Signers may modify this sign using a negative facial expression (accompanied with a specific mouth gesture), for example, to convey an emotion associated with feeling disoriented. These variants are still considered to be modified variants of the lemma EMOTION.

Some signs when modified using facial expression may also modify the movement of the sign. For example, in the sign LOVELY2 (Figure 30), the signer may use a negative facial expression to mean ‘not very nice’. In addition to this, the signer may also modify the path movement of the sign so that the sign is produced slower than usual and the final part of the sign is marked with a repeated downward movement. The movement modification may occur with or without the negative facial expression and still retain its negative meaning. This type of modification has also been observed for other signs such as NICE (ending instead with a single

twisting movement). In each case, the modification observed is best thought of as a type of prosodic modification similar to an English speaker using a sarcastic tone ‘That was niiiice(!)’. In BSL SignBank, such variants are not recognised as separate lexemes. Both variants presented in Figure 30 are considered modified variants of the lemma LOVELY2.



**Figure 30.** Example of prosodic modification: LOVELY2

## 5 Issues in the creation of a sign language dictionary

Thus far we have outlined some of the main issues and problems involved in lemmatising a sign language lexical database. Lexical databases include core lexical signs of the language. Decisions of what signs to include or exclude depend on resources available. As stated in Section 3.2, BSL SignBank is a partial representation of the BSL lexicon based on signs from Brien (1992) and, in particular, the BSL Corpus (Schembri et al. 2014). As more annotation is undertaken on the BSL Corpus, more signs will be added, and further lemmatisation will take place with existing signs in SignBank. This, along with the fact that new additions to SignBank are invited from users, will eventually lead to a resource that is as representative as possible of BSL. At a minimum, each entry within the lexical database should represent the lexeme in citation form. This is not to say that phonological and morphological variants should not be included; instead, their association with other related variants (e.g. the lemma and other variants derived from the lemma) should be acknowledged and consistently represented in the dictionary/lexical database. One way to do this is via double ID glossing within SignBank; one field that serves as head ID gloss (for the lemma, or citation form), and a separate field that gives a unique variant gloss for every entry in the database. We have done this in BSL SignBank via a field called Annotation ID gloss which represents the lemma or citation form (e.g. LOVELY2), and a separate field called ID gloss which represents the particular phonetic, phonological or morphological variant (e.g. the prosodically modified variant of LOVELY2 in Figure 30 may be represented as LOVELY2b where the letter suffix ‘b’ indicates a

related variant of LOVELY2). This allows us to identify both unique lexemes (via Annotation ID gloss, e.g. for computing lexical frequency) and also variants within the database (via ID gloss).

A lexical database serves as a representation of the lexicon of a language for researchers. It can also usefully serve as a dictionary for a variety of users, including learners. However, it is important to recognise that a dictionary may need to serve different functions from a lexical database. For example, the inclusion of entries that are only in citation form might not be appropriate for a dictionary. There are cases where phonetic/phonological/morphological variants might have consistent enough differences in definitions and/or translation equivalents (keywords) to warrant separate entries in the dictionary (although not as separate lexemes), to aid in searchability and linking of particular meanings with particular forms, particularly for learners (e.g. the two variants of LOVELY2 in Figure 30). Additionally, particular variants may be linked to sociolinguistic factors and thus may warrant a separate entry in a dictionary (so that learners are aware of these variants). For example, the sign GREEN in BSL is produced with a B-hand with a movement up the arm and has been identified in a number of regions across the UK. This sign has a phonological variant which is identical except that the movement is down the arm; this variant so far has only been identified in Belfast (Stamp 2013). Although these two variants clearly constitute one lexeme according to our criteria (i.e. they have the same meaning and differ only in one parameter), for the purposes of a dictionary it is useful to have the down-arm variant listed separately since it appears to be linked to a particular sociolinguistic factor (in this case, region).

An additional (and separate) issue in the creation of a sign language dictionary is outlining clear criteria for distinguishing lexical signs (i.e. fixed forms with a specialised meaning) which belong in the lexical database from other types of signs which do not. These types of signs may be fingerspelled sequences or partly-lexicalised constructions - including classifier/depicting constructions, constructed action, and pointing signs - which can differ from lexical signs at the phonological level. Importantly, these types of signs can become lexical signs. However, it is not always straightforward to identify whether a given token encountered in annotation is indeed part of the non-core or non-native lexicon (e.g. a depicting sign or a fingerspelled sequence etc.) or whether there is enough evidence that this token has become lexicalised such that it warrants entry in the lexical database. We have followed principles outlined in Cormier, Schembri and Tyrone (2008) and Cormier, Quinto-Pozos, Sevcikova and Schembri (2012b) in making decisions on the status of tokens as a lexical sign but explicit criteria for determining lexical status for all is necessary – we leave this for future research.

## 6 Conclusion

We have outlined the lemmatisation practices used in the creation of BSL SignBank, a lexical database and dictionary of British Sign Language. In doing so, we have provided some criteria for addressing the *lemma dilemma* for sign languages. The principles we have outlined here should be considered in the creation of any sign language lexical database and ultimately any sign language dictionary and reference grammar. Lexical databases are necessary in the annotation of any sign language data in order to make any kind of generalisations about the lexicon of that sign language. Additionally, the development of large datasets (e.g. corpora) that are consistently annotated in this way at the lexical level will present a significant benefit to those working within the areas of sign language synthesis and automatic recognition of sign languages. These technologies cannot move forward until there are large, annotated, lemmatised corpora to train these systems.

## 7 Acknowledgments

This work was supported by the Economic and Social Research Council of Great Britain (Grants RES-062-23-0825, British Sign Language Corpus Project; ES/K003364/1, Describing sociolinguistic variation in verb directionality in British Sign Language: A corpus-based study; and RES-620-28-0002, Deafness, Cognition and Language Research Centre (DCAL)). We thank Robert Adam, Ramas Rentelis, Katherine Rowley, Rose Stamp, and Bencie Woll for contributing to the development of the criteria outlined here for lemmatisation of BSL SignBank. We thank Sannah Gulamani for acting as model for the figures in this paper. Figures 27 and 28 are from the BSL Corpus (Schembri et al. 2014). We also thank Trevor Johnston for very helpful comments on an earlier draft of this paper. All remaining errors are our own.

---

<sup>1</sup> The first two authors contributed to this paper equally. First author was determined by a coin toss.

<sup>2</sup> Lemmatisation issues for sign languages are likely be more similar to those of spoken languages that lack standard writing systems (see e.g. Chebanne 2010) than those that do have standard orthographies. However, codification of spoken languages is made easier by the fact that there is always at least the International Phonetic Alphabet to start with; sign languages lack an equivalent system.

<sup>3</sup> See section 4.1 for more on sign language phonology.

<sup>4</sup> Unless otherwise noted, use of the term ‘lemmatisation’ in the remainder of this paper will refer to this extended notion of grouping together of phonetic, phonological and morphological variants and distinguishing these from lexical variants.

<sup>5</sup> Johnston and Schembri (1999) do not use the term *lemmatisation* but Johnston (2010) does (p. 118), in the context of ID glossing.

<sup>6</sup> Traditionally, sign language researchers have tended to use contextual glosses for signs or to transcribe signs at a phonetic/phonological level (Miller 2006; Frishberg et al. 2012). Enriching corpora with either or both contextual glosses and phonetic/phonological transcription as an initial step is problematic; both are so slow that they are practically inappropriate as a first step towards enriching corpus data (Johnston 2010, 2014).

<sup>7</sup> As noted by Felbaum (2014), the terms ‘dictionary’ and ‘lexical database’ are often used interchangeably, particularly in the context of digital or electronic lexical and lexicographic

resources. Here, we follow Janssen (2005) in distinguishing dictionaries (lexicographic resources) from lexical databases (resources meant for computational exploitation). There are sign language lexical databases that are research tools but not dictionaries, for example, for Swiss German Sign Language (*Deutschscheizer Gebärdensprache*, DSGS) and for Spanish Sign Language (*Lengua de signos o señas española*, LSE) (Boyes-Braem 2001; Costello and Carreiras 2013).

<sup>8</sup> It is not clear under which circumstances these embellished variants may appear. In contrast to BSL, this type of modification is reported to have an effect on meaning in Auslan, resulting in a semantic shift (Johnston 1989). As a consequence, signs that have been modified in this way are considered as separate lexemes in Auslan.

<sup>9</sup> We recognise that the practice of asking signers to reflect on the meaning of a sign may be problematic. However, these judgments can be verified in future with reference to the BSL Corpus once a sufficient number of tokens have been collected.

<sup>10</sup> It is important to remember that these keywords do not reflect the English mouthings that are used when the sign is produced in connected signing but provide the meanings that are associated with the sign.

## 8 References

### A. Dictionaries

- Brien, D. (ed.). 1992.** *Dictionary of British Sign Language/English*. Boston: Faber & Faber.
- Centre for Tegnsprog. 2008.** 'Ordbog over Dansk Tegnsprog. (the Danish Sign Language Dictionary).
- Fenlon, J., K. Cormier, R. Rentelis, A. Schembri, K. Rowley, R. Adam and B. Woll. 2014a.** *BSL Signbank: A Lexical Database of British Sign Language (First Edition)*. <http://bslsignbank.ucl.ac.uk>. London: Deafness, Cognition and Language Research Centre, University College London.
- McKee, R., D. McKee, S. P. Alexander, L. Pivac and M. Vale. 2011.** 'Online Dictionary of New Zealand Sign Language.
- Stokoe, W., D. Casterline and C. Croneberg. 1965.** *A Dictionary of American Sign Language*. Washington, DC: Gallaudet University Press.

### B. Other Literature

- Baker-Shenk, C. and D. Cokely. 1978.** *American Sign Language: A Teacher's Resource Text on Grammar and Culture*. Silver Spring, MD: T.J. Publishers.
- Battison, R. 1978.** *Lexical Borrowing in American Sign Language*. Silver Spring, MD: Linstock Press.
- Bergman, B. and O. Dahl. 1994.** 'Ideophones in Sign Language? The Place of Reduplication in the Tense-Aspect System of Swedish Sign Language' In Bache, C., Hans Basboll and C.-E. Lindberg (eds), *Tense, Aspect and Action: Empirical and Theoretical Contributions to Language Typology*. Berlin: Mouton de Gruyter, 397-422.
- Boyes-Braem, P. 2001.** 'A Multimedia Bilingual Database for the Lexicon of Swiss German Sign Language.' *Sign Transcription and Database Storage of Sign Information. Special issue of Sign Language and Linguistics* 4.1/2: 133-143.
- Brennan, M. 1992.** 'An Introduction to the Visual World of BSL' In Brien, D. (ed.), *Dictionary of British Sign Language/English*. London: Faber & Faber, 1-133.
- Brennan, M., M. Colville and L. Lawson. 1984.** *Words in Hand: A Structural Analysis of the Signs of British Sign Language*. Edinburgh: Edinburgh British Sign Language Research Project, Moray House College of Education.
- Brentari, D. 1998.** *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.
- Brentari, D. and C. A. Padden. 2001.** 'Native and Foreign Vocabulary in American Sign Language: A Lexicon with Multiple Origins' In Brentari, D. (ed.), *Foreign Vocabulary: A Cross-Linguistic Investigation of Word Formation*. Mahwah, NJ: Lawrence Erlbaum Associates, 87-119.
- Brien, D. and G. Turner. 1994.** 'Lemmas, Dilemmas and Lexicographical Anisomorphism: Presenting Meanings in the First BSL-English Dictionary' In Ahlgren, I., B. Bergman and M. Brennan (eds), *Perspectives on Sign Language Usage*. Durham: The International Sign Linguistics Association.
- Chebanne, A. 2010.** 'The Role of Dictionaries in the Documentation and Codification of African

Languages: The Case of Khoisan.' *Lexicos* 20: 92-108.

- Cormier, K., J. Fenlon, T. Johnston, R. Rentelis, A. Schembri, K. Rowley, R. Adam and B. Woll. 2012a.** 'From Corpus to Lexical Database to Online Dictionary: Issues in Annotation of the BSL Corpus and the Development of BSL Signbank ' In Crasborn, O., E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen and J. Mesch (eds), *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Workshop Part of 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey]*. Paris: ELRA, 7-12.
- Cormier, K., J. Fenlon and A. Schembri. 2014.** Directionality in British Sign Language Is Not Obligatory: The Importance of Corpus Data When Considering "Agreement". *88th Annual Meeting of the Linguistic Society of America, Minneapolis, Minnesota, 2014*.
- Cormier, K., D. Quinto-Pozos, Z. Sevcikova and A. Schembri. 2012b.** 'Lexicalisation and De-Lexicalisation Processes in Sign Languages: Comparing Depicting Constructions and Viewpoint Gestures.' *Language and Communication* 32.4: 329-348.
- Cormier, K., A. Schembri and M. E. Tyrone. 2008.** 'One Hand or Two? Nativisation of Fingerspelling in ASL and BANZSL.' *Sign Language and Linguistics* 11.1: 3-44.
- Cormier, K., A. Schembri and B. Woll. 2013.** 'Pronouns and Pointing in Sign Languages.' *Lingua* 137: 230-247.
- Costello, B. and M. Carreiras. 2013.** LSE Lexicon: Vital Statistics. Paper presented at the *11th Theoretical Issues in Sign Language Research, University College London, 2013*.
- Crasborn, O. 2011.** 'The Other Hand in Sign Language Phonology' In Oostendorp, M. v., C. J. Ewen, E. Hume and K. Rice (eds), *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell, 223-240.
- Felbaum, C. 2014.** 'Large-Scale Lexicography in the Digital Age.' *International Journal of Lexicography*.
- Fenlon, J., A. Schembri, R. Rentelis, D. Vinson and K. Cormier. 2014b.** 'Using Conversational Data to Determine Lexical Frequency in British Sign Language: The Influence of Text Type.' *Lingua* 143: 187-202.
- Frishberg, N., N. Hoiting and D. I. Slobin. 2012.** 'Transcription' In Pfau, R., M. Steinbach and B. Woll (eds), *Sign Language: An International Handbook*. Berlin: Mouton de Gruyter, 1045-1075.
- Janssen, M. 2005.** 'Lexical Vs. Dictionary Databases: Design Choices of the Mordebe System.' *Papers in Computational Lexicography - COMPLEX 2005, Budapest Hungary*.
- Johnson, R. E. and S. K. Liddell. 1984.** 'Structural Diversity in the American Sign Language Lexicon' In Testen, D., V. Mishra and J. Drogo (eds), *Papers from the Parasession on Lexical Semantics*. Chicago, IL: Chicago Linguistics Society, 173-185.
- Johnston, T. 1989.** *The Sign Language of the Australian Deaf Community*. Doctoral dissertation, University of Sydney.
- Johnston, T. 2001a.** 'Nouns and Verbs in Auslan (Australian Sign Language): An Open and Shut Case?'

*Journal of Deaf Studies and Deaf Education* 6.4: 235-257.

- Johnston, T. 2001b.** 'The Lexical Database of Auslan (Australian Sign Language).' *Sign Language & Linguistics* 4.1/2: 145-169.
- Johnston, T. 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, T. 2012.** 'Lexical Frequency in Signed Languages.' *Journal of Deaf Studies and Deaf Education* 17.2: 141-162.
- Johnston, T. 2014.** 'The Reluctant Oracle: Adding Value to, and Extracting Value from, a Signed Language Corpus through Strategic Annotations.' *Corpora* 9.2: 155-189.
- Johnston, T., J. van Roekel and A. Schembri. in press.** 'On the conventionalization of mouth actions in Auslan (Australian Sign language). *Language and Speech*.
- Johnston, T. and A. Schembri. 1999.** 'On Defining Lexeme in a Signed Language.' *Sign Language and Linguistics* 2.2: 115-185.
- Johnston, T. and A. Schembri. 2007.** *Australian Sign Language: An Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press.
- Klima, E. and U. Bellugi. 1979.** *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Konrad, R., T. Hanke, S. König, G. Langer, S. Matthes, R. Nishio and A. Regen. 2012.** 'From Form to Function. A Database Approach to Handle Lexicon Building and Spotting Token Forms in Sign Languages' In Crasborn, O., E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen and J. Mesch (eds), *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*. Language Resources and Evaluation Conference (LREC), Istanbul, May 2012, 87-94.
- Kristoffersen, J. and J. B. Niemelä. 2008.** 'How to Describe Mouth Patterns in the Danish Sign Language Dictionary' In Quadros, R. M. d. (ed.), *Sign Languages: Spinning and Unraveling the Past, Present and Future. Tislr9, Forty Five Papers and Three Posters from the 9th Theoretical Issues in Sign Language Research Conference. Florianopolis, Brazil, December 2006*. Petropolis: Editora Arara Azul, 230-238.
- Kristoffersen, J. and T. Troelsgård. 2012.** 'Electronic Sign Language Dictionaries' In Granger, S. and M. Paquot (eds), *Electronic Lexicography*. Oxford: Oxford University Press, 290-312.
- Lewin, D. and A. C. Schembri. 2011.** 'Mouth Gestures in British Sign Language: A Case Study of Tongue Protrusion in BSL Narratives.' *Sign Language & Linguistics* 14.1: 94-114.
- Liddell, S. K. 1980.** *American Sign Language Syntax*. The Hague: Mouton de Gruyter.
- Liddell, S. K. 2000.** 'Indicating Verbs and Pronouns: Pointing Away from Agreement' In Emmorey, K. and H. Lane (eds), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Lawrence Erlbaum, 303-320.
- Liddell, S. K. 2003.** *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.



- Lillo-Martin, D. and R. P. Meier. 2011.** 'On the Linguistic Status of 'Agreement' in Sign Languages.' *Theoretical Linguistics* 37.3-4: 95-141.
- McDonald, B. H. 1985.** 'Productive and Frozen Lexicon in ASL: An Old Problem Revisited' In Stokoe, W. and V. Volterra (eds), *SLR '83: Proceedings of the 3rd International Symposium on Sign Language Research*. Rome: CNR, 254-259.
- McKee, R. and D. McKee. 2013.** 'Making an Online Dictionary of New Zealand Sign Language.' *Lexicos* 23.1: 500-531.
- Miller, C. 2006.** 'Sign Language: Transcription, Notation, and Writing' In Brown, K. (ed.), *Encyclopedia of Language and Linguistics, 2nd Ed.* Oxford: Elsevier, 353-354.
- Office for National Statistics. 2011.** 2011 Census for England and Wales.
- Padden, C. 1988.** *Interaction of Morphology and Syntax (Outstanding Dissertations in Linguistics, Series IV)* New York: Garland Press.
- Padden, C. A. 1998.** 'The ASL Lexicon.' *Sign Language and Linguistics* 1.1: 39-60.
- Pfau, R., M. Steinbach and B. Woll (eds). 2012.** *Sign Language: An International Handbook*. (Handbooks of Linguistics and Communication Science). Berlin: Mouton de Gruyter.
- Sandler, W. and D. Lillo-Martin. 2006.** *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Schembri, A., J. Fenlon, R. Rentelis and K. Cormier. 2014.** *British Sign Language Corpus Project: A Corpus of Digital Video Data and Annotations of British Sign Language 2008-2014 (Second Edition)*. London: University College London. <http://www.bsllcorpusproject.org>.
- Schembri, A., T. Johnston and D. Goswell. 2006.** 'Name Dropping: Location Variation in Australian Sign Language' In Lucas, C. (ed.), *Multilingualism and Sign Languages: From the Great Plains to Australia* Washington, DC: Gallaudet University Press.
- Schembri, A., D. McKee, R. McKee, T. Johnston, D. Goswell and S. Pivac. 2009.** 'Phonological Variation and Change in Australian and New Zealand Sign Languages: The Location Variable.' *Language Variation and Change* 21.2: 193-231.
- Schembri, A., G. Wigglesworth, T. Johnston, G. Leigh, R. Adam and R. Barker. 2000.** 'The Test Battery for Australian Sign Language Morphology and Syntax Project: Noun-Verb Pairs in Auslan' In Schembri, A., J. Napier, R. Beattie and G. Leigh. (eds), *Deaf Studies, Sydney, 1998: Selected Papers from the Australasian Deaf Studies Research Symposium, Renwick College, August 22-23, 1998*. Sydney: North Rocks Press, 99-118.
- Stamp, R. 2013.** *Sociolinguistic Variation, Language Change and Contact in the British Sign Language (BSL) Lexicon*. PhD dissertation Thesis, University College London.
- Stamp, R., A. Schembri, J. Fenlon, R. Rentelis, B. Woll and K. Cormier. 2014.** 'Lexical Variation and Change in British Sign Language.' *PLoS ONE* 9.4: e94053.
- Sterkenburg, P. G. J. v. (ed.). 2003.** *A Practical Guide to Lexicography*. Amsterdam: John Benjamins.
- Stokoe, W. 1960.** 'Sign Language Structure: An Outline of the Visual Communication Systems of the

American Deaf' *Studies in Linguistics: Occasional Papers (No. 8)*. Buffalo, NY: Dept. of Anthropology and Linguistics, University of Buffalo.

**Supalla, T. and E. Newport. 1978.** 'How Many Seats in a Chair? The Derivation of Nouns and Verbs in American Sign Language' In Siple, P. (ed.), *Understanding Language through Sign Language Research*. New York: Academic Press, 91-132.

**Sutton-Spence, R. and B. Woll. 1999.** *The Linguistics of British Sign Language*. Cambridge: Cambridge University Press.

**Tervoort, B. 1953.** *Structurele Analyse Van Visueel Taalgebruik Binnen Een Groep Dove Kinderen [Structural Analysis of Visual Language Use in a Group of Deaf Children]*, Noord-Hollandsche Uitgevers Maatschappij.

**Troelsgard, T. and J. Kristoffersen. 2008.** 'An Electronic Dictionary of Danish Sign Language' In Quadros, R. M. d. (ed.), *Sign Languages: Spinning and Unraveling the Past, Present and Future. TISLR9, Forty Five Papers and Three Posters from the 9th Theoretical Issues in Sign Language Research Conference. Florianopolis, Brazil, December 2006*. Petropolis: Editora Arara Azul, 652-662.

**Wittenburg, P., H. Brugman, A. Russel, A. Klassmann and H. Sloetjes. 2006.** 'ELAN: A Professional Framework for Multimodality Research' *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. 1556-1559.

**Woll, B. 2001.** 'The Sign That Dares to Speak Its Name: Echo Phonology in British Sign Language (BSL)' In Boyes-Braem, P. and R. Sutton-Spence (eds), *The Hands Are the Head of the Mouth*. Hamburg: Signum Press, 87-98.

**Zwitserslood, I. 2010.** 'Sign Language Lexicography in the Early 21st Century and a Recently Published Dictionary of Sign Language of the Netherlands.' *International Journal of Lexicography* 23.4: 443-476.

**Zwitserslood, I., J. Kristoffersen and T. Troelsgard. 2013.** 'Issues in Sign Language Lexicography' In Jackson, H. (ed.), *The Bloomsbury Companion to Lexicography*. London: Bloomsbury, 259-283.