# Lexical frequency in British Sign Language conversation: A corpus-based approach

KEARSY CORMIER[1], JORDAN FENLON[1], RAMAS RENTELIS[1], & ADAM SCHEMBRI[2]
*University College London, United Kingdom[1] & La Trobe University, Australia[2]*

## 1. INTRODUCTION[1]

An understanding of lexical frequency is important in a variety of subdisciplines within linguistics, including studies of grammaticalisation, language processing, sociolinguistic variation and change as well as first and second language acquisition (e.g., Bybee 2006; Ellis 2002). To date, only three studies on lexical frequency in sign languages have been conducted: on New Zealand Sign Language (NZSL), American Sign Language (ASL), and Australian Sign Language (Auslan). As information regarding lexical frequency in British Sign Language (BSL) has not been readily available, researchers have sought to address this gap by collecting subjective familiarity ratings from 20 deaf signers for 300 BSL lexical signs (Vinson et al. 2008). This paper presents the findings from the first lexical frequency study in BSL based on 24,864 sign tokens collected as part of the BSL Corpus Project (Schembri et al. 2011). The BSL study is also the first lexical frequency study in any sign language to be based entirely on spontaneous conversational data. The results from this study, when compared to previous research into lexical frequency in other signed languages and the familiarity ratings collected by Vinson et al. (2008), suggest that differences in frequency across different sign categories may be attributed largely to situational context and that subjective familiarity ratings may not be a good indication of lexical frequency in sign languages as has been reported for spoken languages (Stadthagen-Gonzalez & Davis 2006).

## 2. BACKGROUND

### *2.1. The British Sign Language Corpus Project*

The lexical frequency study that we report here is based on data from a digital video corpus of British Sign Language (Schembri et al. 2011) which was inspired by developments in the emerging field of documentary linguistics (Woodbury 2003), and particularly in response to the need for sign language documentation to improve sign language description (Schembri 2010). This digital video corpus

[1] The data reported in this paper were collected for the British Sign Language Corpus Project (BSLCP) at University College London, funded by the Economic and Social Research Council UK (RES-620-28-6001), and supplied by the CAVA repository (www.bslcorpusproject.org/cava). The data are copyright. We thank the BSL Corpus Project team for assistance in collecting and analysing the data, and the 249 deaf participants who contributed to the corpus data.

consists of spontaneous and elicited BSL collected from deaf native and near-native signers. The set of participants is 'stratified' (using a non-random quota sampling technique) for gender, region, age, and age of BSL acquisition with 249 signers filmed in 8 key regions across the UK (Belfast, Birmingham, Bristol, Cardiff, Glasgow, London, Manchester and Newcastle) taking part in four tasks: retelling a personal experience narrative, engaging in a 30 minute conversation, participating in an interview on language awareness and attitudes, and responding to a task designed to elicit 102 key lexical items known to vary across the UK. Some of the dataset is partly annotated using ELAN software and all of the video data have been made available online for researchers and/or the wider sign language community (www.bslcorpusproject.org/data).

*2.2. Lexical frequency in signed languages*

There have been only three studies investigating lexical frequency in signed languages. The first project to investigate the distribution of lexical items in a sign language was by McKee & Kennedy (2006) which, drawing on a dataset of 100,000 sign tokens in the Wellington Corpus of NZSL, remains the largest frequency study to date. This was followed by a second, much smaller frequency study by Morford & MacFarlane (2003) who conducted a distributional analysis of 4111 sign tokens in ASL collected from commercially available videotapes of 27 deaf signing individuals. The most recent study was based on 63,436 signs tokens taken from the Auslan archive and presented a cross-linguistic analysis of sign frequency by drawing upon the findings of the previous two studies (Johnston 2011). These three studies drew from similar text types for their analysis. This included data from spontaneous conversation and narrative data, and data involving more formal registers (i.e. interviews and committee meetings). The Auslan frequency study also included sign tokens from narrative retellings and descriptions of a cartoon.[2]

These studies on NZSL, ASL and Auslan report that pointing signs are amongst the most frequent signs in their sign language data. As is the case for signed languages in general, pointing signs generally function as first, second, and third person pronouns, determiners, and locatives. The first person pronoun (i.e. a point to the signer's chest) was the most frequent sign in the NZSL (McKee & Kennedy 2006) and Auslan (Johnston 2011) study and the second most frequent sign in the ASL study. For the ASL study, the non-first pronoun (a category that conflates second and third person pronouns) was the most frequent sign. All three studies also report a relatively low number of function signs compared to lexical frequency studies of English where the frequency of individual function words is higher than that observed in any of the sign language studies discussed here (Leech et al. 2001). However, the three studies suggest that this is not surprising if one considers how signed languages are structured. That is, grammatical functions

[2] Of these three studies, only the Auslan lexical frequency study is based on a corpus which was created to be a large, representative, accessible and machine-readable collection (and thus, a 'corpus' in the strictest sense; cf. McEnery & Wilson 1996).

that are typically marked by functors in English (such as prepositions *in, on,* or *by,* or the conjunctions *but,* or *if*) may instead be marked by modifying signs in space or signalled by accompanying non-manual features. All three studies demonstrate that their respective sign languages are lexically dense, with a relatively high ratio of content signs compared to function signs. The NZSL and Auslan study also demonstrate that a small number of unique signs account for a significant proportion of their data. The NZSL study reports that their top 12 signs account for roughly 20% of their total dataset of 100,000 signs, 116 signs represent 50%, and 665 signs account for 80%. In the Auslan study, the two most frequent signs (PT:PRO1 and G:WELL) account for 8.7% of 55,859[3], the top 10 signs account for 20.5% and the top 100 signs account for 52.8%. To put this into perspective, out of 55,859 sign tokens, 6171 unique signs types were derived. Of the 6171 unique signs types, 100 signs represent nearly 53% of the total dataset of 55,859. Although this seems like a very small number of types accounting for a very large number of tokens, Johnston (2011) makes a point of noting that this is not surprising given that this result is often seen in spoken language frequency studies, as in the distribution of lexical items in the spoken English component of the British National Corpus (Leech et al. 2001).

In the ASL and Auslan study, each sign token was additionally grouped according to sign category: whether it represented a core lexical sign, a pointing sign, a gesture, a classifier sign[4], a fingerspelled sequence, or a name sign (the ASL study includes an additional category for number signs as well). This categorisation of lexical types reflects models of the sign language lexicon, where different major categories (e.g., 'core' lexical signs, 'non-core' pointing and classifier signs and 'non-native' fingerspelled items) represent subcomponents with different morphophonological properties (see Brentari & Padden 2001; Johnston & Schembri 2007). When the distribution of sign category in both studies is considered, it can be seen that signs from the core lexicon represent approximately two thirds of their data (ASL = 73,2%, Auslan = 65%) followed by the second largest category, pointing signs (ASL = 13.8%, Auslan = 12.3%). (Information regarding the distribution of sign category is not fully reported in the NZSL study although core lexical signs also accounted for the bulk of their data.) Classifier signs were the third most frequent category in Auslan and the fourth most frequent category in ASL (ASL = 4.2%, Auslan = 11%). The fourth most frequent category in the Auslan study were gestures (6.9%) which was the lowest occurring category in the ASL study (0.2%). This discrepancy is likely due to differences in glossing practices (i.e. the two studies differ in their definition of a gesture) rather than between the two languages in the extent to which gesture is used. Both the ASL and Auslan studies further divide their data according to text

[3] The Auslan study restricts its attention to signs made with the right hand after no difference between the right and left hand in sign occurrence was observed.

[4] Johnston (2011) uses the term *depicting signs* to refer to classifier signs. Note that we agree that the analysis of these signs as including classifier morphemes is problematic (Schembri 2003), but we have adopted this terminology for ease of comparison with the existing sign language literature.

type (formal interviews, casual conversation, and narratives) so that the distribution of sign category can be viewed across various registers. In doing so, both studies observe an increase across text types in the proportion of pointing signs beginning with formal interviews (ASL = 5.8%, Auslan = 7.4%), narrative data (ASL = 13.4%, Auslan = 15%) and casual conversation (ASL = 17.3%, Auslan = 16.1%) and an increase in the proportion of classifier signs beginning with formal interviews (ASL = 0.9%, Auslan =1.6%), casual conversation (ASL = 1.1%, Auslan = 7.3%) and finally the narratives (ASL = 17.7%, Auslan = 21.4%). These studies highlight, as has been observed with spoken language frequency data (e.g., Johansson 1985), the potential for text type to influence the distribution of sign category.

### *2.3. Familiarity ratings of British Sign Language signs*

As lexical frequency data for signed languages was previously not readily available, researchers have attempted to address this problem by obtaining frequency information via other means. Vinson et al. (2008) report on a study that collected subjective familiarity ratings for 300 lexical signs in BSL. Here, the researchers sought to benefit from a possible link between familiarity and lexical frequency reported in Balota et al. (2004). The aim was to provide crucial information for psycholinguists hoping to control for lexical frequency in language processing experiments until objective frequency data for BSL became available (although Balota et al. report that, of the two, frequency is by far the stronger predictor of lexical processing effects).

Twenty deaf participants, the majority of which reported BSL as their preferred everyday language, were asked to watch 300 BSL signs and indicate on a scale of 1-7 how often they saw the sign (1 being 'I have never seen the sign before' and 7 being 'I see this sign everyday'). Items rated as most familiar were all concepts suspected to be used in everyday conversation (e.g. WORK (M = 6.90), EAT (M = 6.80), and WHAT (M = 6.80) and items rated as least familiar included signs that were likely to be known only in a specific region such as BASINGSTOKE (M = 1.95).

Johnston (2011) compares the familiarity ratings collected in Vinson et al. (2008) to his Auslan frequency data and reports that familiarity ratings may not be a reliable indicator of a sign's frequency[5]. Of the 300 lexical signs selected for the familiarity study, 157 occur in the Auslan frequency data, 26 (8.7%) of which occur in the top 100 ranked fully lexical signs and 57 (19%) in the top 300. Additionally, 127 of the 157 signs returned a high familiarity rating in Vinson et al. (i.e. a rating of 5 or higher) but only 18 (14.2%) of these highly familiar signs appear in the top 100 and only 39 (29.9%) in the top 300 in the Auslan data. This comparison of two closely related sign language varieties suggests, as a precursor to our comparison here, that the relationship between familiarity and lexical

[5] This comparison between familiarity in BSL and frequency in Auslan is appropriate because BSL and Auslan are historically related, generally considered to be dialects of the same language (McKee & Kennedy 2000).

frequency may not be as straightforward as previously thought. Furthermore, Johnston (2011) stresses that these results reflect only a comparison between the 300 lexical items selected in the BSL familiarity study and the core lexical signs occurring in his data, since only core lexical signs were included in the BSL familiarity study. If his rankings were modified to reflect all the other sign categories that occur naturally in everyday signed conversation (e.g. gestures, classifier signs, pointing signs) then the relationship between familiarity and frequency may be further weakened.

## 3. LEXICAL FREQUENCY IN BRITISH SIGN LANGUAGE

In this paper, we report on a study of lexical frequency in BSL, based on 24,864 sign tokens from the BSL Corpus conversation data: approximately 500 signs each from 50 participants, 'stratified' (non-randomly selected to fit quotas) for age, region, gender, and age of BSL acquisition.

### *3.1. Methods*

All 24,864 sign tokens were grouped according to the following sign categories: signs from the core lexicon (also known as 'lexical signs'), pointing signs, classifier signs, gestures[6], sequences of fingerspelling, buoys (see Liddell 2003) and name signs. Signs from the core lexicon are signs which are highly conventionalised in form and meaning across contexts (Johnston & Schembri 1999). Each unique sign from the core lexicon of BSL was assigned its own ID gloss (often, an equivalent English word that represented the 'best fit' with the sign's meaning) which is associated with that particular variant and all its related, lexicogrammatically-modified realisations - providing this does not change a sign's meaning in which case a separate ID gloss is required. This ID gloss was then entered into the project's lexical database along with a range of key English equivalents that relate to its meaning in BSL. Pointing signs include pronominals, locatives, determiners and possessives. The classification of pointing signs into these grammatical functions was, however, one of the most significant challenges in this study. If a given token's function was ambiguous between two possibilities, both possible functions were included in the gloss (e.g. PT:PRO3/LOC) but, as was often the case, when its function in a particular context was even more ambiguous, it was labelled as a pointing sign alone with no additional gloss (e.g. PT). Classifier signs were further divided according to whether they represented classifier constructions of motion, location, visual-geometric description, or handling. The category of gesture was very broad because it includes a wide variety of communicative actions, from gestures that serve a discourse function (e.g. the use of G:WELL to facilitate the flow of conversation), to those that

[6]We adopt this terminology here as a working hypothesis about the different kinds of meaningful unit in sign languages, but we note that there are good reasons for abandoning a sharp distinction between 'gesture' and 'sign' (Kendon 2008).

encourage lexical retrieval (e.g. a form glossed as G:ERM), to sequences of constructed action where the signer enacts an action of a referent directly. With regard to fingerspelling, we make a distinction between signs that are derived from fingerspelling and are part of the core lexicon (e.g. the signs MOTHER, FATHER and GOVERNMENT which are based on the initial manual letter for each English word) and any fingerspelled sequence where the signer fully fingerspells a word (e.g. FS:TWININGS). The fingerspelling category listed here refers to the latter exclusively.

All the annotation for this project was carried out using ELAN, a multimedia software package that allows the precise time alignment of annotations to corresponding media files (http://www.lat-mpi.eu/tools/elan/). This is particularly advantageous as annotations and the primary digital video data can then be referred to again quickly when reviewing glossing practices. Additionally, ELAN allows for the data to be counted and exported to Microsoft Excel for further quantitative and statistical analysis.

*3.2. Results and discussion*

Preliminary results based on 24,864 sign tokens indicate some similarities to the previous sign frequency studies. In Table 1 below, the top 10 most frequent signs out of 2507 different signs that occur in our data are listed.

**Table 1**
Top 10 most frequent signs in BSL conversations

| Rank | ID gloss | Total | % | % (cumulative) |
|---|---|---|---|---|
| 1 | PT:PRO1 | 1717 | 6.9% | 6.9% |
| 2 | G:WELL | 1360 | 5.5% | 12.4% |
| 3 | PT:PRO3 | 955 | 3.8% | 16.2% |
| 4 | PT | 789 | 3.2% | 19.4% |
| 5 | GOOD | 477 | 1.9% | 21.3% |
| 6 | PT:PRO2 | 408 | 1.6% | 23.0% |
| 7 | PT:DET | 394 | 1.6% | 24.5% |
| 8 | PT:LOC | 346 | 1.4% | 25.9% |
| 9 | SAME | 253 | 1.0% | 26.9% |
| 10 | RIGHT | 228 | 0.9% | 27.9% |

Table 1 indicates that the most frequent sign in our data is the first person pronoun (PT:PRO1) which accounts for 6.9% of the 24,864 tokens. The next most frequent sign is one that we have classed as a gesture (G:WELL) which accounts for 5.5% of our total. This sign can be described as a palm-up gesture that is often used as a discourse marker (similarly to English 'well') and can convey a variety of meanings often via a change in non-manual features. When viewed by sign category, Table 1 indicates that pointing signs in general are amongst the most frequent signs in our data occupying 6 of the top 10 places (along with 3 lexical

signs and 1 gesture). This observation is consistent with the findings from the other sign language frequency studies. Five of the top 10 signs in the NZSL study are pointing signs, as are 4 in the ASL study and 3 in the Auslan study. The differences in the number of pointing signs between these studies are also likely to be due to glossing practices (e.g. the ASL and Auslan study both conflate second and third person pronouns and we have an additional category for ambiguous points that, strictly speaking, may not represent a unique pointing sign in itself). Additionally, the Auslan study reports, as we do here, that the second most frequent sign in their data is the same gesture glossed here as G:WELL. An apparently identical form is listed as the 14th most frequent sign in the ASL study.

In the rightmost column of Table 1, a set of percentages that represents the cumulative frequency of the top 10 signs is provided. For example, when the total number of tokens glossed as PT:PRO1 and G:WELL are added together, this combined set represents 12.4% of the 24,864 tokens. That is, over one tenth of our data consists of only two lexical items. The final cell in this column also indicates that the top 10 signs, when added together, form 27.9% of our total data. Further analysis of the data beyond the top 10 signs shows that the top 100 signs (out of a total of 2507 different signs observed in the data) account for 57.2% of our data. The fact that a large proportion of the data is represented by a small number of lexical items has also been reported for the sign frequency studies discussed here as well as for frequency in spoken language corpora (e.g., Leech et al. 2001).

When the 24,684 total is divided according to sign category, our results indicate that 62.0% (n=15372) of the dataset consists of signs from the core lexicon. The next two largest categories are pointing signs (22.9%, n = 5697) and gestures (i.e. gesture-like signs and sequences of enactment or constructed action) (8.7%, n=2174). The remaining number of tokens consists of fingerspelled signs (2.5%, n=659), classifier constructions (i.e., classifiers of motion and location, size and shape classifiers, and handling classifiers, 2.3%, n=566), and sign names, some of which are fingerspelled (1.0%, n=261). The distribution by sign category is provided in Table 2 below together with distributional data from the ASL and Auslan studies.

Preliminary comparisons with the ASL and Auslan data reveal some interesting similarities and differences. Firstly, the frequency of pointing signs is much higher in the BSL conversational data (22.9%) when compared to the ASL and Auslan data overall (13.8% and 12.3% respectively). Note, however, that both Morford & MacFarlane (2003) for ASL and Johnston (2011) for Auslan found that the frequency of pointing signs was greater in casual signing contexts (i.e. in conversation) when compared to more formal contexts (i.e. interviews). Additionally, the frequency of gesture tokens is greater in the BSL data (8.7%) than in ASL (0.2%) and Auslan (6.5%), but this may reflect coding differences, as explained above. The frequency of BSL classifier constructions is similar to the ASL data (4.2%) but lower than that reported for Auslan (11%). This may reflect the larger proportion of elicited narrative text-types in the Auslan corpus, some of which were specifically selected to elicit classifier signs of motion, location and handling. Some of the differences in the frequency of different categories of signs

across the studies may be due to differences in genre. For example, the fact that the BSL lexical frequency study described here consists solely of free conversation suggests that pointing signs and gestures may be more typical in conversation than other genre types. Indeed, when this comparative analysis is restricted to the conversational data from the Auslan and ASL study (described in both as the 'casual' genre), one finds a closer similarity between the three languages in the distribution of sign categories. Both ASL and Auslan report a slight increase in the proportion of pointing signs (ASL = 17.3%, Auslan = 16.1%) and a slight decrease in the proportion of classifier signs (ASL = 1.1%, Auslan = 7.3%) when the data is restricted to their casual text types. It remains to be seen if these differences will persist in frequency studies based on larger datasets.

**Table 2**
Distribution of sign categories in BSL, Auslan and ASL

| Sign category | BSL (n =24,864) | ASL (n = 4111) | Auslan (n = 63,436) |
|---|---|---|---|
| 'Core' lexical signs | 62.0% | 73.2% | 65.0% |
| Fingerspelling | 2.5% | 6.4% | 5.0% |
| Pointing signs | 22.9% | 13.8% | 12.3% |
| Classifier signs | 2.3% | 4.2% | 11.0% |
| Gestures | 8.7% | 0.2% | 6.5% |
| Name signs | 1.1% | 2.3% | 0.2% |
| Buoys | 0.5% | n/a | n/a |
| Unknown | 0.1% | n/a | n/a |

When compared to spoken language frequency data collected for English (Leech et al. 2001), all three sign language studies report a low number of functional signs amongst the most frequent items in the language. An examination of the top 100 most frequent signs in the BSL data reveals similar findings. Based on the ID gloss given to each unique sign, we observe 22 functors amongst the top 100 most frequent signs in our data. This is a similar number to the 24 function signs occurring in the top 100 of the Auslan study and much lower than that reported for spoken language frequency studies: Johnston (2011) points out that the spoken component of the British National Corpus lists 56 function words within the top 100 most frequent words in English.

A preliminary comparative analysis with the familiarity ratings collected by Vinson et al. (2008) suggests that familiarity may not be a reliable indicator of a sign's actual frequency. We compared each of the 100 signs rated as most familiar from Vinson et al. (out of a possible 300) to its occurrence in our frequency data; the 100 most familiar signs all received a mean familiarity rating of 5.9 or higher on the scale (7 being highly familiar and 1 meaning not familiar at all). The

results reported here are based only on the category of core lexical signs in our frequency data (i.e. excluding pointing signs, gestures, classifier signs, etc.). In other words, we restrict our attention here to 15,409 sign tokens all representing signs from the core lexicon in which we observe 1535 unique signs.

Out of the 100 signs rated as most highly familiar in Vinson et al. (2008), only 84 occur in our data and only 3 occur in our 50 most frequent signs which represent 36% of 15,409 sign tokens. We find only 8 highly familiar signs from Vinson et al. (2008) that occur within our 100 most frequent signs (representing 49.4% of 15,409 tokens) and 21 highly familiar signs occur in our 300 most frequent signs (representing 74.1% of 15,409 tokens). Thus, it appears that signs rated as highly familiar do not often occur within the core lexical signs that make up a significant proportion of our data. As Johnston (2011) notes, the link between frequency and familiarity may be further weakened when all sign categories are included (e.g. MUST, the third most familiar sign in Vinson et al.'s data drops from a ranking of 44 amongst lexical signs only to 62 when all sign categories are included).

This initial study of lexical frequency in BSL provides much needed evidence for BSL researchers about frequency which can be used to design experiments about BSL processing. It has also helped further our understanding of distribution of sign categories (and unique sign types) within BSL, across related (BSL, Auslan, NZSL) and unrelated (ASL) sign languages, and across signed and spoken languages. Such comparisons both within and across language modalities are crucial for better understanding of language processing, acquisition, variation and change in language in general, and also for theory building.

## REFERENCES


Balota, D.A., M.J. Cortese, S.D. Sergent-Marshall, D.H. Spieler & M.J. Yap. 2004. Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General* 133. 283-316.

Brentari, Diane & Carol A. Padden. 2001. Native and foreign vocabulary in American Sign Language: A lexicon with multiple origins. In Diane Brentari (ed.), *Foreign vocabulary: A cross-linguistic investigation of word formation*, 87-119. Mahwah, NJ: Lawrence Erlbaum Associates.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82. 711-733.

Ellis, Nick C. 2002. Frequency effects in language processing. *Studies in Second Language Acquisition* 24. 143-188.

Johansson, Stig. 1985. Some observations on word frequencies in three corpora of present-day English texts. *I.T.L. Review of Applied Linguistics (67-68)*. 117-126.

Johnston, Trevor & Adam Schembri. 1999. On defining lexeme in a signed language. *Sign Language and Linguistics* 2. 115-185.

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

Johnston, Trevor. 2011. Lexical frequency in signed languages. *Journal of Deaf Studies and Deaf Education*.

Kendon, Adam. 2008. Some reflections on the relationship between 'gesture' and 'sign'. *Gesture*. 348-366.

Leech, Geoffrey, Paul Rayson & Andrew Wilson. 2001. *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.

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

McEnery, Tony & Andrew Wilson. 1996. *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

McKee, David & Graeme Kennedy. 2000. Lexical comparisons of signs from American, Australian, British and New Zealand Sign Languages. In Karen Emmorey & Harlan Lane (eds.), *The Signs of Language revisited: An anthology to honor Ursula Bellugi and Edward Klima*, 49-76. Mahwah, NJ: Lawrence Erlbaum Associates.

McKee, David & G. Kennedy. 2006. The distribution of signs in New Zealand Sign Language. *Sign Language Studies* 6. 373-390.

Morford, Jill & James MacFarlane. 2003. Frequency characteristics of American Sign Language. *Sign Language Studies* 3. 213-225.

Schembri, Adam. 2003. Rethinking "classifiers" in signed languages. In Karen Emmorey (ed.), *Perspectives on classifier constructions in sign languages*, 3-34. Mahwah, NJ: Lawrence Erlbaum Associates.

Schembri, Adam. 2010. Documenting sign languages. In Peter Austin (ed.), *Language documentation and description volume 7: Lectures in language documentation and description*, 105-143. London: School of African and Oriental Studies.

Schembri, Adam, Jordan Fenlon, Ramas Rentelis & Kearsy Cormier. 2011. British Sign Language Corpus Project: A corpus of digital video data of British Sign Language 2008-2009 (First Edition). London: University College London. http://www.bslcorpusproject.org.

Stadthagen-Gonzalez, Hans & Colin J. Davis. 2006. The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods* 38. 598-605.

Vinson, David P., Kearsy Cormier, Tanya Denmark, Adam Schembri & Gabriella Vigliocco. 2008. The British Sign Language (BSL) norms for age of acquisition, familiarity and iconicity. *Behavior Research Methods* 40. 1079-1087.

Woodbury, Anthony. 2003. Defining language documentation. In Peter Austin (ed.), *Language documentation and description*, Vol. 1, 35-51. London: School of Oriental and African Studies.