As of January 12, 2009, the downloaded file has 304,817 rows. Each field is separated by a backslash: \
We strongly recommend that you email Rachel Channon at rachel.channon@uconn.edu to ask her to put you on a list to receive updates. Any information about how you use the database, and bibliographic references to papers based on it would be appreciated.
We welcome questions, which will help us improve the documentation.

1 Introduction and acknowledgements

SignTyp is funded by the National Science Foundation (grant BCS0544944). The primary goal of this first phase was to develop a prototype of a crosslinguistic database containing phonetic-phonological encoding of signs in a variety of sign languages, using datasets that had already been coded by various researchers. Setting a goal comparable to, for example, CHILDES (MacWhinney 2000), SignTyp provided a home for existing datasets from different linguists. We also developed a uniform coding system into which the coding systems of our various sources have been translated. The primary focus of this system is to notate form properties of sign that are potentially distinctive or phonological, so that researchers can analyze these potentially distinctive values, and determine which elements are actually non-distinctive or phonetic and which are truly phonological. Multiple languages will also allow the determination of which characteristics are cross-linguistically distinctive.

The database currently includes 9 different language sets, for six different sign languages, from four annotating groups or individuals: ASL (3 different time periods), NGT (Sign Language of the Netherlands), Korean Sign Language, Japanese Sign Language, New Zealand Sign Language, and Finnish Sign language (two different time periods). We would like to thank the individuals who provided us with these datasets: Daisuke Hara, Lorna Rozelle and the SignPhon group (Harry van der Hulst, Onno Crasborn, Els van der Kooij, Marja Blees and others – see Crasborn, van der Hulst and van der Kooij 2001). The Hara and Rozelle data were the datasets they used in their dissertations (Hara 2003 and Rozelle 2003), and the SignPhon data was used for dissertations by Crasborn (2001) and van der Kooij (2002). An additional ASL dataset was provided by Rachel Channon based on the Long (1918) ASL dictionary.

The datasets have been integrated with a flexible database architecture developed by Rachel Channon in collaboration with Robert Cochran, a web programmer, and David Graff from the Linguistic Data Consortium at the University of Pennsylvania. To the best of our knowledge, this is the largest phonetically annotated database of signs that is publicly available. All interested parties (e.g., researchers, developers of practical tools) can access the SignTyp data on the web where they can download the data to use within their own database software (for example, Access, Excel, Fourth Dimension etc.).

A (non-exhaustive) list of possible uses of SignTyp:

Theoretical use:

- Establish comparative type and frequency information about *inventories* of building blocks of signs comparable to the work by Ladefoged and Maddieson (1996) on spoken languages (van der Hulst and Channon 2008).

- Establish comparative information about *the simultaneous and sequential structure* of signs (see van der Hulst and Channon 2008, Channon and van der Hulst to appear, van der Hulst to appear; Channon and van der Hulst in prep a). For example, Channon and van der Hulst (to appear) argue that dynamic features such as path shape and directions are required.

- Establish comparative information about *specific properties* of types of signs or 'strata' of sign language lexicons (such as monomorphemic core words, fingerspelled words, classifier constructions, compounds, etc. as in Brentari and Padden (2001).

- Develop and test theories about *dependencies* (co-occurrence restrictions, redundancies) between properties of signs. For example, hooked hand flexion tends to co-occur more often with a single extended finger (and the others closed), whereas curved handshapes most often occur in handshapes with all four fingers curved (Channon and van der Hulst in prep).

- Establish comparative information about *iconicity*, e.g. how do languages differ with respect to iconic encoding of comparable semantic/referential properties? For example, do sign languages more often encode handling or shape characteristics of objects?

- Establish generalizations about *processes of emergence* and *diachronic change* in sign languages

- Establish comparative information about differences between signed and spoken languages (as in van der Hulst and Channon, to appear, where we compare the segmental structure of signed and spoken languages using data from CELEX (Baayen, Piepenbrock & Gulikers 1995) and SignTyp).

- Establish base lines for understanding what degree of similarity in sign languages indicates historic relatedness as opposed to similarities caused by iconicity, as in Parkhurst & Parkhurst 2003.

- establish a base line of citation form characteristics to compare to non-citation (naturalistic) signing

Practical uses:

- Development and testing of *notation and writing systems* for sign languages (see van der Hulst and Channon, in press).

- Development of *teaching materials* for teachers of sign languages and linguistics.

- Development of *lexicographic materials* for sign languages especially endangered ones.

- Development of *sign recognition and synthesis software*.

To promote and demonstrate the use of SignTyp we have worked on several of these types of usage, with four papers either in press or in preparation.

2 Using SignTyp to find universals and differences across sign languages

One of our most important theoretical goals was (and is) to find universals across sign languages. For that purpose, it can even be said that the more disparate the sources the better, because that means that any **similarities** found are very robust, if those differences can surface in spite of variations in underlying theoretical assumptions in the different datasets. For example, we are able to show (Channon and van der Hulst to appear), that signs rarely have more than two locations, and most signs have only one location. A second paper presented at our conference and in preparation for publication now will show that there is little difference between languages in terms of the most common handshape characteristics. For example, cross-linguistically, the most common value for finger extension is for the fingers to be extended; the most common number of extended fingers is 4, followed by 1; the most common extended finger is the index; hooked handshapes are more often 1 hooked finger (with the other fingers closed) rather than 4 hooked fingers, while curved handshapes have the opposite pattern of more often 4 curved fingers rather than 1 curved and the others closed. (Some of these findings have been reported before on a narrower set of data (for example van der Kooij 2002).

It should be noted however, that SignTyp as currently constructed should be used with caution if you are looking to find **differences** between sign languages or groups of sign languages. Because of these disparate sources, it can be difficult to decide whether differences are attributable to differences in assumptions and annotations or are real differences between languages. A simple example of this problem is that our data from Hara for JSL and ASL is based on the notation system used in Stokoe, Casterline and Croneberg (1965), and does not list any signs as made on the weak hand, because they considered such signs to be double handshape (double dez) signs made in neutral space. The other sources considered some signs where the weak hand was contacted as being signs made on the weak hand. So before using this database, we recommend that you become familiar with the types of data available for each source (by creating pivot tables or similar frequency tables of the database sorted by source).

3 The SignTyp database structure and coding system

SignTyp is a physical database, a coding system and a database structure. It can be used with any database software. It uses almost no abbreviations or special symbols, but rather ordinary English words. For some values, such as handshape descriptions, this is a closed set of English words, for others, such as glosses, any English word is possible.

The data structure for SignTyp is the same structure used in phonological feature trees (directed graphs)[1]. However, instead of being a graphic representation, it is tabular, because graphical representations are not suitable for analytic work. (Of course, SignTyp includes a much broader range of data than a phonological tree, but this does not affect the structure.)

Consider the directed graph (tree) representations of location for ASL BEAUTIFUL and FLOWER[2], shown in Figure 1, rotated so that the root is oriented left. Table 1 shows corresponding tabular trees.

Figure 1 Partial feature trees for BEAUTIFUL and FLOWER rotated

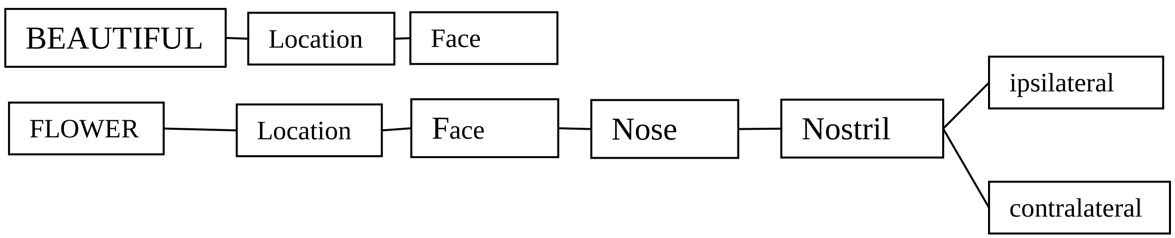


Table 1 SignTyp records for BEAUTIFUL (SignID 1) and FLOWER (SignID 2)

SignID	Stage	Heading	Field Values Detail			
			Level 1	Level 2	Level 3	Level 4
1	1	Location	Face			
2	1	Location	Face	Nose	Nostril	Ipsilateral
2	2	Location	Face	Nose	Nostril	Contralateral

The graphic tree uses lines to show the relationship between the nodes: *Face* is a dependent of *Location*. In SignTyp, the root node is the leftmost column and the terminal node is the rightmost column. Columns to the right are dependents of columns to the left. The left to right organization of the columns and the repetition of all non-terminal nodes on each row substitute for the lines in the graphic tree form.

A graphic tree has one additional convention: sister nodes are temporally ordered left to right. In the tabular tree, the Stage value orders the data, so both representations provide the same information for dependency and linear order.

The SignTyp structure means that related fields are automatically and naturally sorted together, and pivot tables or similar data summaries will always result in like material being kept together. For example, location information will sort separately from orientation.

---

[1] SignTyp can also be seen as an attribute-value structure in which a value can itself be an attribute (Scobbie 1997).

[2] BEAUTIFUL: fingers moves in circular gesture over the face as a whole. FLOWER: hand contacts ipsilateral and contralateral nostrils.

Similar material will be aligned. BEAUTIFUL's terminal node is *face*, while FLOWER's terminal nodes are the contralateral and ipsilateral *nostrils*. Both FLOWER and BEAUTIFUL share a *face* node, which shows up in the same column for both signs. In a traditional row-column structure, *face* for BEAUTIFUL and *nostril* for FLOWER would probably be put in the same column, which would miss a generalization. Data with more detail will have more nodes, but all the data can be compared on any shared higher nodes. This also means that data from different databases that code different depth of detail can be united in SignTyp (after normalization of labels). For example, data source A might treat all locations on the chest as a single location while data source B might distinguish upper chest, mid chest, lower chest and so on. Both databases can be compared at Detail Level 1 where they would both have a value of chest, even though DataSource A has no further information available for any sign, while DataSource B has much more information available.

With an ordinary relational database, once the database administrator has set up the database structure, and the data structure, the analyst can add more rows at any time without needing to ask the administrator to change anything. But if the analyst wanted to add a new field, *iconic location*, the database administrator (and probably the website designer/administrator) would need to review, approve and make this change. This can be a problem in a research database, which is much more likely to add or change fields than a business database, because the analytic discovery process frequently means that new fields will be added.

In SignTyp *iconic location* is not a new column, but just a new possible value for a record. No database changes would be required. New kinds of data are simply new rows of data with different node names. This structure goes a long way to freeing the researcher from dependence on the database and website administrators.

SignTyp has almost 12,000 signs, and over 700,000 rows of information (when expanded to include handshape information from the separate handshape table). Table 2 shows the sources for SignTyp and Table 3 shows the kinds of information available for each data set.

Table 2 SignTyp dataset characteristics

Sign Language	Original Coder	Based on:	Signs
American Sign Language	Rachel Channon	Long 1918	1549
	Daisuke Hara	Stokoe , Casterline & Croneberg 1965	1891
	Lorna Rozelle	Tennant & Brown 1998	656
Japanese Sign Language	Daisuke Hara	Japan Institute for Sign Language Studies 1997	2516
Korean Sign Language	Lorna Rozelle	Kim 1993	614
Finnish Sign Language	Lorna Rozelle	Hirn 1910	354
	Lorna Rozelle	Liitto 1998	609
New Zealand Sign Language	Lorna Rozelle	Kennedy, Graeme, et al. 1999	688
Sign Language of the Netherlands	SignPhon group	Videotapes and other materials collected in the 1990s	3052
Total signs			**11929**

Table 3 Rows of Information by Source and Type for SignTyp

Data Set	Action	Handshape	Location	Orientation	Contact	Other	Total
Hara ASL	3350	50429	4085		1415	16913	76192
Hara JSL	3839	65107	5240		1350	22135	97671

Long ASL	2890	33700	3524		2799	22143	65056
Rozelle ASL		19993	1850			6440	28283
Rozelle KSL		19943	1676			6509	28128
Rozelle NZSL		19452	1775			6521	27748
Rozelle SVK		17442	1670			6813	25925
Rozelle VSVK		10572	936			3807	15315
SignPhon	12675	182513	21333	23494	3329	101049	344393
Total	**22754**	**419151**	**42089**	**23494**	**8893**	**126334**	**708711**

* Other includes morphology, semantics, sources, nonmanuals, & prosody & miscellaneous

3.1 *Example of coding*

The Long (1918) dictionary includes the following description on page 28:

My and Mine.- Press open hand, palm against breast to indicate possession.

Table 4 and Table 5 show how this information, and knowledge about the source, was used to create database entries for SignID Long028_060.

Table 4 Signs Table (general information about the sign)

Major Category	Heading 1/2	Value
morphology	compound status	noncompound
morphology	gloss/English	mine
morphology	gloss/English	my
source	dataset	Long
source	SIL language code	ase
source	source code	Long 1918

Table 5 Sign Features Table (stage specific information about the sign)

Major Category	**Heading 1/2/3**	**Value 1/2**
Stage 1: entity described is sign		
hand	articulator symmetry	one hand
hand	number of hands in stage	one hand
Stage 1: entity described is strong hand		
action	Repetition	no repetition
contact	Contacting Place/hand side/frontbackness	palm
hand posture	crossing	no crossing
hand posture	finger group/closed fingers	no closed fingers
hand posture	finger group/less extended	no secondary extension

hand posture	finger group/most extended	IMRP (index, middle, ring, pinky)
hand posture	finger group posture/less extended	no secondary extension
hand posture	finger group posture/most extended	extended
hand posture	fingers contacted by thumb	no contact
hand posture	number of nonclosed fingers	4
hand posture	overall shape	extended
hand posture	thumb contact	no contact
hand posture	thumb posture	extended in same plane as palm
hand posture	unity of handshape	unitary
location	major location	body/torso
location	verticality/body	torso/chest

4 Fields in the database

Note that *SignID* plus *Stage* does NOT provide unique identifiers for each record in SignTyp, but *SignID* **is** a unique identifier in the file SignIndexTable. All fields have potential lengths of 255, except for the memo field. All fields are alphanumeric.

The SignTyp coding system was devised to be extremely user-friendly. With only minimal training, an analyst should be able to understand and remember the values used, which are ordinary English words with almost no codes. Two exceptions are that the Ethnologue language code is included, and in fields that refer to the digits of the hand, each digit is represented as a single letter: TIMRP = thumb, index, middle, ring, pinky). There should be no need to consult a table or documentation to determine the meaning of some value.

Table 6 SignTyp fields

Field	**Explanation**
SignID	Partial key (most signs will have more than one record with the same *SignID* and *Stage*) Will always have some value. The first part of the key usually indicates the source database. The second part is usually related to the original ids used in the source database, or to page numbers or other sequencing information. For example, SignPhon_0986 is from the SignPhon database and had an original id of 986. Long028_070 is based on a definition from page 28 of the Long dictionary, and 070 is a sequence number such that all definitions on the page will sort in the same order that they occur on the page. Notice that these sequence numbers will have gaps: for example, Long028_070 could be followed by Long028-073. Although this field is called SignID, it is actually just a kind of partial record id – for example handshape records will have an id identifying a particular handshape, not a sign. Note that the name *SignId* is a misnomer for some of the tables. For example the "SignID" for the *LanguageList* is the *LanguageID*. If you decide to create separate tables for the *LanguageList* records, you would of course want to rename this column to LanguageID for that table. This also holds for the *Actors, Handshapes,* and *Sources* tables. (The name should be simply ID, but that is too likely to be a reserved name in some database implementation, so SignID was used.)
Stage	Subkey (most signs will have more than one record with the same SignID and *Stage*) Has the placeholder value of 0 for those records for which a *Stage* is meaningless, (such as glosses, and other records in *Signs, Actors*, *Handshapes*, *languages*, *Sources* tables*)*. Has non-zero values for *SignFeatures*.

Field	**Explanation**
SignTypTable	Which table this data should be sent to. Currently includes: *Actors, Handshapes, LanguageList, Sources, Signs, SignFeatures*. Additional tables may be added later. This column indicates which table the record belongs to. If you create a relational database with these records, you can choose to create separate tables following the names in this field, or you can leave these records in a single master table. The new tables will look just like this table except that this field (*SignTypTable*) won't be included. If additional tables are added later, the plan is for each table to have the same format. At present, there are two additional tables with a different format: a parent index table *SignIndex* that will have one record for each SignID, and a second index table *SignIDStageIndexTable*. This second table is a convenience to make it easier to select records based on their SignID and stage value.
EntityDescribed	A small set of values that lists which entity is being described: hand, eye, mouth, the sign as a whole, strong hand, weak hand etc.
MajorCategory	A major categorization of the data. It include values such as *Sign, Location, Handshape, Action, Orientation, Semantic Information* and other large categories used in describing signs.
FieldHeading	This is the equivalent of a column heading in the traditional column and row spreadsheet, and describes the type of data that will be seen in the *FieldValues*. For example, if the *MajorCategory* is *Action*, a *FieldHeading* might be *handshape change*.
FieldSubHead1, FieldSubHead2	These are the equivalent of column subheadings in the traditional column and row spreadsheet, and describe the type of data that will be seen in the *FieldValues*. For example, if the *MajorCategory* is *Action*, and *FieldHeading* is *Gloss*, then *FieldSubHeading1* might be *Dutch*.
FieldValue1	This column gives the primary value for the specific *FieldHeading* plus subheadings, *MajorCategory*, entity and table given in the previous columns. For example if the table is *SignFeatures*, the *EntityDescribed* is the *strong hand*, *MajorCategory* is *orientation*, the *FieldHeading* is *palm*, then *FieldValue1* could be *upward*, *downward*, *frontward*, etc. In some cases, *FieldValue1* is the foreign key for other tables.
FieldValue2, FieldValue3, FieldValue4, FieldValue5, FieldValue6	These columns provide additional information about the value provided in *FieldValue1*. In many cases, these fields will have nothing except an underscore value (_).
Memo1	Unlimited length field Some values are inserted here instead of in the FieldValues. Usually this was because the original data was a memo field or remark of some kind. Some of these will probably be moved to the FieldValues column eventually, but any type of record which is at least sometimes greater than 255 characters will appear here. This is the only field which can be longer than 255 characters. Some sample uses would be for remarks or in copying the entire text of a description of a sign from a dictionary.

The SignIndex Table has one record for each SignID. The purpose of this file is to act as an index for the records in the tables Sign and SignFeatures

Table 7 SignIndex Table fields

Field	Explanation
SignID	Alphanumeric key Will always have some value. The first part of the key indicates the source database. The second part is usually related to the original ids used in the source database, or to page numbers or other sequencing information. For example, SignPhon_0986 is from the SignPhon database and had an original id of 986. Long028-070 is based on a definition from page 28 of the Long dictionary, and 070 is a sequence number such that all definitions on the page will sort in the same order that they occur on the page. Notice that these sequence numbers will have gaps: for example, Long028-070 could be followed by Long028-073.
ConvDatabase	This indicates the source of the converted data. Because data sources vary in their level of detail and in the available fields, it is often important to know what was the original data source. Every SignID will always have one and only one ConvDatabase.
LanguageCode	This is the language code from the SIL Ethnologue (http://www.ethnologue.com/) which provides a unique language code for human languages. Every SignID will always have one and only one LanguageCode.

The SignIDStageIndex Table has one record for each SignID and stage. The purpose of this file is to act as an additional index for the records in the tables Sign and SignFeatures.

Table 8 SignIDStageIndex Table fields

Field	Explanation
SignID	Alphanumeric key Will always have some value. The first part of the key indicates the source database. The second part is usually related to the original ids used in the source database, or to page numbers or other sequencing information. For example, SignPhon_0986 is from the SignPhon database and had an original id of 986. Long028-070 is based on a definition from page 28 of the Long dictionary, and 070 is a sequence number such that all definitions on the page will sort in the same order that they occur on the page. Notice that these sequence numbers will have gaps: for example, Long028-070 could be followed by Long028-073.
Stage	Subkey (most signs will have more than one record with the same SignID and *Stage*) Has the placeholder value of 0 for those records for which a *Stage* is meaningless, (such as glosses, and other records in *Signs*, *Actors*, *Handshapes*, *languages*, *Sources* tables*)*. Has non-zero values for *SignFeatures*.
ConvDatabase	This indicates the source of the converted data. Because data sources vary in their level of detail and in the available fields, it is often important to know what the original data source was. Every SignID will always have one and only one ConvDatabase.
LanguageCode	This is the language code from the SIL Ethnologue (http://www.ethnologue.com/) which provides a unique language code for human languages. Every SignID will always have one and only one LanguageCode.

5 Table relationships:

The foreign key is located in the SignID for Signs and SignFeatures.

Table 9 Table relationships

Table	**Child of**	**Sample key values**
SignIndex	None. Parent table	SignPhon_1501 SignPhon_1046 SignPhon_1045
Signs	SignIndex The SignIndex table will have one record for each SignID; the Signs table will have at least one, and usually many, records with a given SignID.	SignPhon_1501 SignPhon_1046 SignPhon_1045
SignFeatures	SignIndex The records will have the same SignID as the records in Signs with the addition of a Stage value. This table has a one-to-many relationship with the SignIndex table. The SignIndex table will have one record for a SignID; the SignFeatures table will usually have many records for each SignID. Each record in SignFeatures will also have a Stage, but the SignID plus Stage is not a unique identifier in the Sign_Features table. There can be many records with the same SignID and Stage.	SignPhon_1501 SignPhon_1046 SignPhon_1045

6 Tables that use foreign keys in *Signs* or *SignFeatures.*

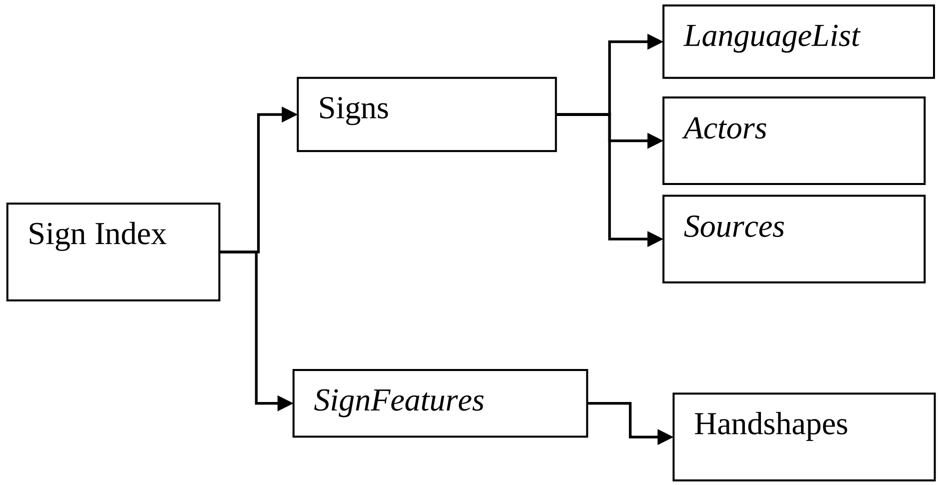
Note that foreign keys are always located in *FieldValue1* for these child tables.

Table 10 Foreign Keys

Table	**Child of**	**Foreign key is located in table**	**The records that provide the foreign id will have the following values**			**Sample key values**
			Entity Described	***Major Category***	**FieldHeading**	
Handshapes	*SignFeatures*	*SignFeatures*	Varies (strong hand or weak hand)	Handshape	handshape code	RozelleHS_0047 LongHS_IMRP IM crossed LongHS_IM extended crossed HaraASLHS_r SignPhonHS_d17
LanguageList	*Signs*	*Signs*	sign	source	SIL language code	ase dse fse jsl
Sources	*Signs*	*Signs*	Sign	Source	source code	Kennedy 1999 Japan Institute 1997
Actors	*Signs*	*Signs*	Sign	Source	signer code	SignPhonSigner_1 SignPhonSigner_5 SignPhonSigner_0

Note: Orphans, or child records without parent records, can occur for *LanguageList*, *Handshapes*, *Sources* and *Actors*. For example, there are many records in *LanguageList* which are never referenced, because there are no signs for that language. The language records have been entered now, with the idea that future databases may have records for these languages.
Orphans should not occur for *Signs* or *SignFeatures*.
Keys and other information are not case sensitive.

Figure 2 table relationships



Every record in *SignIndex* will have at least one record in *Signs*. In most cases, a record in *SignIndex* will have multiple records in *Signs* and multiple records in *SignFeatures*. All three tables are linked by the *SignID*. The *SignID* is a unique identifier for *SignIndex* but NOT for *Signs* or *SignFeatures*.

Some records in *Signs* will have a foreign key in *FieldValue1*. This key will link to a key in the *SignID* field in *LanguageList* or *Actors* or *Sources*.

Some records in *SignFeatures* will have a foreign key in *FieldValue1*. This key will link to a key in the *SignID* field in *Handshapes*. For example the *SignFeatures FieldValue1* might be *SignPhonHS_b04*. Then there will be (usually) multiple records in *Handshapes* with a *SignID* of *SignPhonHS_b04*. If the *SignFeatures* record *FieldHeading* says *handshape code*, then it should always be true that there is a *Handshapes* record or records with a *SignID* that equals the value in *FieldValue1*. However, it can happen that there are *Handshapes* records which never occur in *SignFeatures*.

The situation with *LanguageList*, *Actors* and *Sources* is the same as for the *SignFeatures* and Handshape tables.
Some records in *Signs* will have a foreign key in *FieldValue1*. This key will link to a key in the *SignID* field in *LanguageList*, *Actors* or *Sources*. For example the *Signs FieldValue1* might be *ase*. Then there will be (usually) multiple records in *LanguageList* with a *SignID* of *ase*. If the *Signs* record *FieldHeading* says SIL language code, then it should always be true that there is a *LanguageList* record or records with a *SignID* that equals the value in *FieldValue1*. However, it can happen that there are *LanguageList* records which never occur in *Signs*. The same is true of *Actors* and *Sources*. Every *SignID* will have one and only one link to *LanguageList*. Usually, it should also have zero or one link to *Actors* and *Sources*, but there could be multiple links.

7 General remarks about the data structure

It is important to recognize that many fields can have multiple rows of similar information. An obvious example is glosses. Many signs have more than one gloss.

Another example is neutral space locations. These locations can have three dimension records, for height, laterality and front-back dimensions. Many records will have only the height dimension; some signs will have no dimension records.

8 Notes on the handshape variables

The handshape variables used in SignTyp are relatively easy to determine from descriptions or pictures, and do not require knowledge of a sign or a language. These variables were chosen because they were simple to determine from picture sources, and appear to cover all of the possible distinctive values for handshapes in citation form signs. They are not intended to describe minor differences in aperture or flexion. Their use of course does not preclude other more theoretically based handshape variables, or the addition of variables showing finer distinctions. For example, a variable such as selected fingers could be added, but this would require knowledge of both the specific sign and the language constraints. Table 11 lists the values used to describe handshapes with some explanations and examples.

Most common handshapes are relatively simple. In the most common cases, all of the fingers do the same thing. But there are a number of more complex handshapes. With the exception of a very few handshapes used in a few signs, where each finger is extended to a different degree (see the stacking variable discussion below), it seems that all other handshapes can be handled with a maximum of three finger groups and a description of the thumb. The first group is the most extended fingers, the second group is those fingers that are not closed but are less extended than the most extended fingers, and the last group is any closed fingers. In almost all cases, only one or two of these groups would be needed for any particular handshape. For example, the ASL 5 handshape, with all fingers extended and spread will have a most extended finger group but no secondary extension and no closed fingers. The ASL 1 handshape, with the index extended from the fist, will have a most extended finger group (the index) and a closed finger group (MRP). The ASL F handshape has the middle, ring and pinky extended with the index curved to contact the thumb. In this handshape, the most extended finger group would be MRP, with a posture of *extended*, the less extended finger group would be I, the index finger, with a posture of *curved*, and no closed finger group. Handshapes like this would also be labeled as *complex* for the overall shape variable. An example of a complex handshape that would require using all three finger groups would be ASL K, where the index is the most extended finger, the middle is bent (less extended) and the ring and pinky are closed.

An important advantage of this variable set (as opposed to a variable naming the handshape as a whole) is that in some cases it is not possible to determine what all of the fingers are doing. For example a description might say that the index is extended, but nothing is said about the posture of the other fingers or the thumb. In such a case, a value can be given for the most extended finger group and its posture, but less extended fingers, closed fingers, and thumb can be left blank or marked as unknown. This is also valuable in cases where the non-selected fingers may have various closed or slightly open postures without affecting the meaning of the sign.

Notice also, that while in most cases, most extended fingers can be equated with selected fingers, this is not always true. Especially (and probably only) when there is a less extended finger set, the selected finger set could be the less extended finger set, or in some cases, selected fingers may be ambiguous. The ASL D handshape is probably an example of ambiguous selection – the index is extended, but the other three fingers are not closed, but form a circle with the thumb. Thus a researcher interested in the selected finger notion is probably safe to consider the most extended fingers of handshapes as the selected fingers **if all other fingers are closed**, but would need to make individual determinations for those handshapes with less extended fingers.

8.1 Extended, curved, bent, hooked, closed fingers

An extended finger is one in which all of the knuckles are straight or nearly so. A curved finger has all of the knuckles slightly flexed, but not so flexed that the finger looks closed. A bent finger has only the biggest knuckle (the one between the palm and the finger) flexed. The degree of flexion can vary from approximately 45 degrees to a 90 degree angle. (At the highest degree of flexion (90 degrees), the finger would be considered closed if the other two knuckles were also similarly flexed.) A hooked finger is the "opposite" of a bent finger, because it has the two smallest knuckles flexed and the large knuckle is extended. Finally, a closed finger has all of the knuckles highly flexed.

These values are intended to be slightly impressionistic. For example, there are cases where a closed finger might not be fully flexed, but the finger or the hand looks closed. There are also cases where whether a finger is considered extended or bent, or extended or curved could be debated. That is, some fingers might be considered extended even if there is a slight amount of flexion at one or more knuckles.

Table 11 Handshape variables

Variable name	**Explanation**	**Value Examples**
finger group – most extended group	lists the fingers that are most extended	I M MR no extended fingers variable extension of digits
finger group – less extended group	lists the fingers for those complex handshapes where some fingers are extended and other fingers are not as extended but not closed	I M MR No secondary extension
finger group – closed fingers	Lists any closed fingers	IMRP I IM No closed fingers
finger group posture – most extended	posture for the fingers that are most extended	Extended Curved Bent Hooked no extended fingers
finger group posture – less extended	posture for the less extended finger group (if any)	Curved Bent Hooked no extended fingers
number of nonclosed fingers	total number from 0 to 4 of non-closed fingers	0 1 2 3

		4
overall shape	looks at all fingers except closed fingers (unless all fingers are closed) to decide on an overall shape of extended, bent, curved, hooked, closed, complex or ring-flattened (a type of bent handshape with thumb contact).	Extended Curved Bent Hooked Complex Ring flattened closed
spreading		Unspread Spread One finger
stacking	Used when there is variable extension of fingers. The variable is taken from Brentari's 1998 model, and means a set of fingers which vary in their degree of extension. This is especially valuable for the rare handshapes where all four fingers have varying degrees of extension. For example, in the initial hand posture of ASL FEW, the large knuckle of each finger varies, from straight for the index finger, slightly bend for the middle, more bent for the ring, and closed for the pinky. Another example is the ASL K handshapes, where the index and middle both have straight middle and small knuckles, but the large knuckle of the index finger is extended or only slightly flexed, while the large knuckle of the middle finger is definitely flexed.	ASL K handshape – index and middle extended with middle less extended
crossing	Whether any digit (including the thumb) crosses under one or more fingers	Crossed fingers Crossed thumb No crossing
crossing digits	The specific digits that are involved in crossing	IM crossed Thumb under IM
thumb contact	What the thumb contacts	Between finger(s) Next to index No contact On most extended fingers On second most extended fingers Palm contact no finger contact Under finger(s) on closed fingers
thumb contact type	How the thumb contacts the fingers or Palm	Crossing contact Folded into palm Next to index No contact

		Restraining
thumb posture	What the thumb looks like	Opposed Extended in same plane as palm Closed hooked
unity of handshape	If all of the fingers and the thumb have the same posture then it is labeled as *unitary fingers and thumb*. If all of the fingers but not the thumb have the same posture it is labeled as *unitary fingers*. Otherwise it is considered *non-unitary*. ASL 5 is a unitary fingers and thumb handshape, ASL B (thumb folded into palm) is a unitary fingers handshape, and ASL 1 is a non-unitary handshape.	Unitary fingers and thumb Unitary fingers Non-unitary

9 Some remarks on using the database: SQL, Excel, Access

We have used MS Access to work with the database. In addition, we have generally treated the Signs and SignFeatures tables as a single table. We have found it helpful to use self joins and left joins in working with the database. A self join means that a table is related to itself as TableA and TableAliasA. A left join means that instead of selecting only those records in two tables that match on the keys, all of the records in one table are selected, and only those records in the other table that match are selected.

For example, if you wanted to ask a question about the relationship between locations and handshapes, you might set up a query that selects all location records from the database, and a second query to locate all handshape records. If you only want those records where there is both a location and a handshape for a given sign and stage, you can use a regular join. But if you want to see all locations including those that do not have an associated handshape, you would want to use a left join. (In this example, it is worth pointing out that some of the databases did not repeat the handshape or location in the second stage, if there was no change.)

We have used Access as the storage facility for the database. Access queries can be helpful in selecting the particular information desired, but final analysis seems to work better in Excel. Data in Access can be linked to Excel using either tables or pivot tables. Tables are more useful when the repeating information is needed; pivot tables work well to provide frequency counts.

10 Pivot tables for the data

Copies of pivot tables for the database (with no actual link to the database) are included as a separate MS Excel 2003 file. They are in an Excel file because some of the pivots are very long. Because there is no actual link to the database, they are not functioning pivot tables, but they can be used to get a general feeling for what the data looks like. They are intended only to provide a general understanding of the database, and will not always be updated for minor changes.

The tables show different levels of detail for the database. Some examples are given here to show how to read them.

Table 12 shows all of the tables in SignTyp, and the number of records for each table. (Note that there are only 300,000 records because the handshape table is not expanded – if each sign had all of its handshape fields, there would be more than 700,000 records.)

Table 12 Pivot table for Tables in SignTyp

SignTypTable	Total
actors	16
handshapes	9087
LanguageList	144
SignFeatures	193796
signs	101773
sources	41
Grand Total	304857

Table 13 shows the kinds of entities within each table. As can be seen, the SignFeatures table has the most possible entities. A record in the SignFeatures table could be describing some characteristic of the cheek, the eye, the strong hand, the weak hand, the sign as a whole, and so on.

Table 13 SignTyp tables and entities

SignTypTable	Entitydescribed	Total
actors	actor	16
handshapes	hand	9087
LanguageList	language list	2
	languages	142
SignFeatures	body	7
	cheek	99
	chin	1
	eye	2514
	eyebrows	226
	eyelid	13
	eyes	2

	face	517
	head	26
	head or eye	6
	jaw	6
	lip	333
	mouth	2354
	mouth corner	135
	nose	66
	shoulders	4
	sign	64297
	sound	3
	strong hand	111971
	teeth row	53
	tongue	123
	unknown actor	3
	unknown articulator	8
	weak hand	11029
signs	sign	101773
sources	source	41
Grand Total		304857

Table 14 shows a small portion of a pivot table for all fields except the memo field. As it happens, this fragment also points out two things that the analyst should be aware of. Row 8 shows that there are 348 occurrences of a location value *lower face*.

First, it's important to be aware that this does not equate directly with signs, although it is probably close, because one sign might have more than one occurrence of the same location. This might occur because the sign is a compound, or because the value was repeated for a second stage. Thus in order to obtain a location per sign count, one might want to look for all cases where the same location occurs in two succeeding stages, and eliminate from consideration all duplicates.

Second, as mentioned above, certain datasets have more or less detail. In this case, the Hara data listed only whether the hand was at the lower face, mid face or upper face. This means that it is incorrect to claim that the two largest lower face locations are the mouth and the lower face as a whole.

All of the lower face values can be totaled, and compared to the values for mid face or upper face. But the Hara data must be excluded when looking at any more detailed locations than lower face as a whole.

Table 14 Sample rows for pivot showing all fields in SignTyp

SignTyp Table	Entity described	Major Category	Field Heading 1	Field Sub head 1	Field Sub head 2	Field Value 1	Field Value 2	Field Value 3	Field Value 4	Field Value 5	Field Value 6	Total
Sign features	Strong hand	Location	verticality	body	_	_	_	_	_	_	_	24
						arm	_	_	_	_	_	13
							elbow	_	_	_	_	30
							lower	_	_	_	_	89
							upper	_	_	_	_	55
						head	_	_	_	_	_	128
							face	_	_	_	_	195
								lower face	_	_	_	348
									chin	_	_	235
										under	_	82
									mouth	_	_	387
										corner	_	47
										lips	_	18
											lower	10
											upper	4
										teeth	_	10
										tongue	_	4

11 References

Baayen, R. Harald, Richard Piepenbrock and Leon Gulikers. 1995. The CELEX Lexical Database (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

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

Brentari, Diane. 1998. A prosodic model of sign language. Cambridge MA: MIT.

Channon, Rachel and Harry van der Hulst 2008, to appear. Dynamic features, number of places and path movement in sign languages in Hulst & Channon (eds.), The Phonetics and Phonology in Sign Languages. Ishara Press.

Channon, Rachel and Harry van der Hulst. in preparation. Handshape characteristics and handshape variables in the SignTyp database.

Crasborn, Onno, Harry van der Hulst and Els van der Kooij. 2001. SignPhon: A phonological database for sign languages. Sign Language and Linguistics 4: 1/2, pp. 215-228.

Crasborn, Onno. 2001. Phonetic Implementation of Phonological Categories in Sign Language of the Netherlands. Doctoral dissertation, Leiden University, The Netherlands.

Hara, Daisuke. 2003. A Complexity-Based Approach to the Syllable Formation in Sign Language. Doctoral dissertation, University of Chicago.

Hirn, D. F. 1910. *Kuuromykkäin Viittomakieli*. Helsinki: Osakeyhtiö Weilin and Göös Aktiebolag.

Japan Institute for Sign Language Studies (ed.). 1997. The Japanese-Japanese Sign Language Dictionary. Tokyo, Japan: Federation of the Deaf.

Kennedy, Graeme et al. 1998. A dictionary of New Zealand Sign Language. Auckland: Auckland University Press - xlii, 755 p.

Kim, Seung-Gook. 1993. *Standard Sign Language Dictionary*.

Ladefoged, Peter and Ian Maddieson. 1996. The Sounds of the World's Languages. Oxford: Blackwell.

Liitto, Kuurojen. 1998. *Suomalaisen viittomakielen perussanakirja*. Helsinki, Finland: Kuurojen Litto.

Long, J. Schuyler. 1918. *The Sign Language: A Manual of Signs*. Washington DC: Gallaudet College.

MacWhinney, Brian. 2000. The CHILDES project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Parkhurst, Stephen and Dianne Parkhurst. 2003. "Lexical comparisons of signed languages and the effects of iconicity." Work Papers of the Summer Institute of Linguistics, University of North Dakota Session 47. http://www.und.edu/dept/linguistics/wp/2003.htm.

Rozelle, Lorna. 2003. The Structure of Sign Language Lexicons: Inventory and Distribution of Handshape and Location. Doctoral dissertation, University of Washington.

Scobbie, J. 1997. *Autosegmental Representation in a Declarative Constraint-based Framework*. New York & London: Garland Press. [Original dissertation: Attribute Value Phonology. Edinburgh.

Stokoe, William C. & Casterline, Dorothy C. & Croneberg, Carl G. 1965. A dictionary of American Sign Language on linguistic principles. Washington: Gallaudet College Press

Tennant, Richard A & Brown, Marianne Gluszak. 1998. The American Sign Language handshape dictionary. Washington, DC: Gallaudet University Press - 408 p.

van der Hulst, Harry and Rachel Channon. 2008. Annual Report to the National Science Foundation for grant 0544944.

van der Hulst, Harry. to appear. What is Phonology? in Hulst & Channon (eds.), The Phonetics and Phonology in Sign Languages. Ishara Press.

van der Kooij, Els. 2002. Phonological Categories in Sign Language of the Netherlands: The Role of Phonetic Implementation and Iconicity. Doctoral dissertation, Leiden University, The Netherlands.