

Draft

Developing a corpus to capture the acquisition of Hong Kong Sign Language: Data information and transcription conventions

Gladys Tang, Scholastica Lam, Cat H-M Fung
Centre for Sign Linguistics and Deaf Studies
Chinese University of Hong Kong

Table of Contents

1	Introduction	2
1.1	<i>Background of the Child HKSL Corpus</i>	2
1.2	<i>Deaf Community in Hong Kong</i>	2
1.3	<i>Child Background</i>	2
1.4	<i>Video Sessions (First batch of data)</i>	3
2	Data Transcription System	4
2.1	<i>Structure of Lines</i>	4
2.2	<i>Signs in the Main Line, Glosses 1 and 2</i>	5
2.2.1	Signs	5
2.2.2	Special coding	9
2.2.3	Comments on signs	10
2.2.4	Utterance properties glossed in the Main Line	11
2.3	<i>Tagging</i>	12
2.3.1	Grammatical Categories	12
2.3.2	Functional Tagging	12
2.3.3	Complex Tagging	12
2.3.4	Alternative Tagging	12
2.3.5	Morphological Tagging	13

Appendix 1

1 Introduction

1.1 *Background of the Child HKSL Corpus*

The Child HKSL Corpus has been set up since 2002 as part of a research project entitled “Development of Hong Kong Sign Language by Deaf Children” supported by Hong Kong Research Grant Council #4278/01H. The major goal of building up this corpus is to facilitate sign language acquisition research which is understudied in the field. This corpus is a joint effort from a team of deaf and hearing researchers at the Centre for Sign Linguistics and Deaf Studies: Felix Sze, Kenny Chu, Brenda Yu, Anita Yu, Phippen Wong, Fion Wong, Joe Mak, Cat Fung and Ada Lau. Other colleagues at the Centre also helped at various stages of its development. They are: Jafi Lee, Denise Chan, Jeff Choi, Jenny Lam and Michael Choi, Prudence Lau and Andrew Lam.

The corpus started out with collecting data from one deaf child, CC, who was identified with the help of the Audiological Services Section, the Education and Manpower Bureau, Hong Kong SAR Government. After a number of meetings with the parents, the parents agreed to join the project. Since then the research team visited the child once per week, except when the child was sick.

1.2 *Deaf Community in Hong Kong*

The deaf community in Hong Kong, like those in other parts of the world, is heterogeneous. Different deaf people have different degrees of hearing loss.

In US, only 10% of deaf children are born to deaf parents. Though no similar statistics are reported in Hong Kong, the deaf researchers have the impression that deaf children were rarely born to deaf parents. Most of their deaf peers were born to hearing parents who choose not to sign, but speak.

Given the strong oralist mentality (which highlights speech-reading and residual hearing in deaf education), most deaf children are not exposed to the conventional HKSL. Even if they were born of deaf parents, they may not receive sign language input as HKSL is considered as an inferior code of communication in the community, if not gesture. Deaf schools also encourage deaf children to speak. Therefore the identification of deaf families is difficult in this pro-oralist environment.

1.3 *Child Background*

The corpus is set up with the language data produced by a deaf child named CC. His biodata is as follows:

Name of Child	CC
Birth of Child	2000/5/8
Hearing Loss	bilateral conductive hearing loss; date of test 17/11/03 (age 3;3)
Sex of Child	M
Parents	Non-native HKSL signers

The child was brought up in a largely hearing environment. He was living with working deaf parents and was taken care of by a hearing grandmother and a helper. He spent the day in a special child care centre which was basically adopting an oral approach in educating deaf children.

After receiving the consent from the deaf parents, the research team started to visit CC on a weekly basis, except when the child was sick or when the child needed to join some preschool activities. During each session, at least one trained deaf researcher interacted with the child using various activities to elicit data from the child. These activities ranged from picture naming, story telling, free conversation, and occasionally experimental elicitation. HKSL is the major medium of

Draft

communication in the corpus.

Given the fact that deaf community is heterogeneous, the research team is going to collect data from deaf children who are not born to deaf parents, but hearing parents in order to provide a fuller picture of the child acquisition in signed languages. Data on other children will be added to the corpus phase by phase.

1.4 Video Sessions (First batch of data)

No.	Date of Video-taping	File Name	Age	Location of Video-taping	Time duration
1	2002/6/8	CC010927.cha	1;9.27	Home	1:00:47
2	2002/7/13	CC011021.cha	1;10.21	Home	1:09:38
3	2002/7/27	CC011122.cha	1;11.22	Home	1:24:10
4	2002/8/31	CC020026.cha	2;0.26	Home	1:18.07
5	2002/9/14	CC020109.cha	2;1.9	Home	1:00.44
6	2002/10/5	CC020200.cha	2;2.0	Home	0:46.26
7	2002/11/30	CC020325.cha	2;3.25	Home	0:59:36
8	2002/12/28	CC020423.cha	2;4.23	Home	1:00:06
9	2003/1/18	CC020513.cha	2;5.13	Home	0:44:21
10	2003/2/22	CC020617.cha	2;6.17	Home	1:02:01
11	2003/3/24	CC020719.cha	2;7.19	Home	0:55:10
12	2003/4/23	CC020818.cha	2;8.18	Home	0:59:56
13	2003/6/3	CC020929.cha	2;9.29	Home	0:47:37
14	2003/6/14	CC021009.cha	2;10.9	Home	0:55:14
15	2003/7/26	CC021121.cha	2;11.21	Home, Centre	1:00:09
16	2003/8/18	CC030013.cha	3;0.13	Home	0:59:54
17	2003/9/20	CC030115.cha	3;1.15	Home	0:59:20
18	2003/10/29	CC030224.cha	3;2.24	Home	0:58:00
19	2003/12/4	CC030329.cha	3;3.29	Centre	1:02:00
20	2003/12/18	CC030413.cha	3;4.13	Centre	0:58:05
21	2004/1/28	CC030523.cha	3;5.23	Centre	0;59:39
22	2004/3/4	CC030628.cha	3;6.28	Home	0:59:39
23	2004/3/18	CC030713.cha	3;7.13	Home	0:54:23
24	2004/4/24	CC030819.cha	3;8.19	Centre	0:59:19
25	2004/5/29	CC030924.cha	3;9.24	Centre	1:01:31
26	2004/7/3	CC031028.cha	3;10.28	Home	1:00:00
27	2004/7/31	CC031126.cha	3;11.26	Home	0:53:14
28	2004/8/28	CC040023.cha	4;0.23	Home	0:59:28
29	2004/10/2	CC040127.cha	4;1.27	Home	1:00:07
30	2004/10/30	CC040225.cha	4;2.25	Centre	1:01:13
31	2004/11/27	CC040322.cha	4;3.22	Centre	0:40:35
32	2004/12/18	CC040413.cha	4;4.13	Centre	0:50:20
33	2005/1/8	CC040503.cha	4;5.3	Centre	1:00:47
34	2005/2/26	CC040621.cha	4;6.21	Centre	0:58:41

2 Data Transcription System

This section describes the transcription system used in the HKSL child corpus. We basically follow MacWhinney's (2000) transcription system in *The CHILDES Project: Tools for Analyzing Talk* in transcribing and tagging the HKSL data. Readers are advised to read the CHAT system before reading this manual. Special symbols are added to mark language-specific properties of HKSL. This section is organized as follows. First, we will show how production of signs is transcribed in different lines (Section 2.1). Detailed description on the transcription system in the main line is given in Section 2.2. Finally we will describe the tagging system in Section 2.3.

2.1 Structure of Lines

The data was first transcribed with ELAN, software developed by Max Planck Institute of Psycholinguistics at Nijmegen, later converted to CLAN with the help of MacWhinney's team. Production of Signs of each participant is transcribed with four lines: (i) main line, (ii) gloss 1 (e.g. xgl1@BRE 'gloss 1 of Brenda'), (iii) gloss 2 (e.g. xgl2@BRE 'gloss 2 of Brenda') and (iv) morpheme tier (%mor). Example 1 provides an overview on the lines where transcription are placed for a participant named Brenda (*BRE):

Example 1 – Structure of Lines:

Lines	HKSL Data	Information noted
*BRE	IX_1 ELDER_SISTER IX_3 ELDER_SISTER SAME.	Sign Utterance
%xgl1@BRE	IX_1 > IX_3 SAME.	Individual Signs
%xgl2@BRE	ELDER_SISTER ELDER_SISTER	Individual Signs
%mor	n:pro IX_1 n ELDER_SISTER n:pro IX_3 adj SAME .	Tagging

Note: the symbol '>' means the previous sign continues to hold in the signing space. This is a phenomenon commonly observed in signing.

The first three lines (i.e. main line, gloss 1 and gloss 2) are transcription of HKSL signs and they are time-aligned. The difference between the main line and glosses 1 and 2 is that the former shows the sign utterances and the latter are individual signs. Signs are generally transcribed into gloss 1. When simultaneity occurs, signs would be transcribed into gloss 2, as indicated by the bolded words in the following example:

Example 2 – Glosses 1 and 2

```
%xgl1@BRE IX_1 BUY CANDY HAVE. gesture [= get someone's attention]
%xgl2@BRE IX_2 CANDY HAVE_NOT_HAVE?
```

Note that classifier predicates and two-handed signs are placed in gloss 1. But if a classifier predicate co-occurs with a sign, signs will be placed in glosses 1 and 2 (see the bolded words in example 3 below):

Example 3 – Classifier Predicates and Signs

```
%xgl1@BRE FEMALE walk+CL_sem [= a girl walks] > walk+CL_sem [= a man walks towards the girl]
%xgl2@BRE MALE
```

Transcription on individual signs is necessary for a number of reasons. First, by aligning the individual signs with the time, researchers can examine the signs one by one. This is particularly useful as signs are transcribed in English glosses rather than phonetic coding. Second, two different signs can be articulated simultaneously in signed languages. Two signs may be overlapped partially or they begin at the same time, as illustrated with the bolded words in the following examples:

Example 4 – Signs are overlapped partially:

%xgl1@BRE IX_1 > IX_3 SAME.
%xgl2@BRE ELDER_SISTER ELDER_SISTER

Example 5 – Signs are articulated at the same time (i.e. same Begin Time):

%xgl1@BRE fire_a_gun+CL_sass:gun [= fire a gun] move+CL_sass:bullet [= a bullet moves forward to a person]
%xgl2@BRE be_located+CL_sem:thief [= a thief] CL_sem:thief

%xgl1@BRE DIE FINISH walk+CL_sem [= walk] CANNOT
%xgl2@BRE die+CL_sem [= die]

By showing how signs are overlapped in glosses 1 and 2, researchers can examine simultaneity more closely.

Finally, a morpheme tier indicates the categories of the signs. In addition to grammatical categories like noun, verbs, etc., we also tag categories which are unique to signed languages, e.g. IX = index signs, CL = classifier predicates. This tier is a dependent tier which is linked to the main line. It allows researchers to examine the morphological and syntactic information as decomposed from the utterance tier.

2.2 Signs in the Main Line, Glosses 1 and 2

This section describes the transcription system of the signs (and gestures) in the main line and glosses 1 and 2. We glossed the interaction in HKSL among deaf researchers, hearing researchers and the child. Adult-adult interaction that is out of CC's sight and signing towards CC's sister are excluded as our main goal is to explore CC's language development.

2.2.1 Signs

2.2.1.1 Basic Signs

- Signs are glossed in English words with the closest meanings of the signs.
- Following the sign language literature, signs are glossed in Capital Letter (e.g. LIKE).
- When a sign is expressed by more than one English word, underscore is used to link them up.
- When two or more signs with the different meanings can only be translated with one English word, we use _A, _B to denote different sign types (e.g. FULL_A 'maximum quantity', FULL_B 'cannot take more food').
- When two more signs with the same meaning can only be expressed with one English words, the different sign types are represented with _1, _2, etc. (e.g. IGNORE_1, IGNORE_2).
- But different sign types are not indicated in the substituted signs (glossed within the symbols [*]) because these signs are identified on the basis of meaning and it is hard to tell whether the substituted sign is _1 form or the _2 form. See Section 2.2.2.2.2 below.
- When two or more English words suit the sign meaning, we choose the more general one in glossing (e.g. We chose SEE where SEE, WATCH, LOOK-AT may suit a sign meaning).

2.2.1.2 Index Signs

Index signs are signs in the form of pointing. Seven types of index signs are defined in the present corpus:

- Personal pronouns:
 - IX_1 (first person singular personal pronoun)
 - IX_2 (second person singular personal pronoun)
 - IX_3 (third person singular personal pronoun)
 - IX_1p (first person plural personal pronoun)
 - IX_2p (second person plural personal pronoun)

IX_3p (third person plural personal pronoun)

- Index signs which refer to objects¹
IX_obj [= description of the object];
e.g. IX_obj [= book]
- Index signs which refer to location
IX_loc [= description of the location];
e.g. IX_loc [= kitchen]
- Index signs which refer to a classifier handshape
e.g. IX_CL
- Index signs which refer to a gesture
e.g. IX_ges
- Index signs which refer to a previously established loci in the signing space
IX-a [= bridge]
Note: The ‘-a’ indicates the imagined locations of the IX. This marker is also used for marking actual locations. See Section 2.2.1.3.1 below.
- Index signs which refer to an event or a command
IX_evt [= description of the event];
e.g. IX_evt [= younger sister cries]

IX_cmd [= command]
e.g. IX_cmd [= go there]
- Index signs with unknown referents:
IX [= xxx]

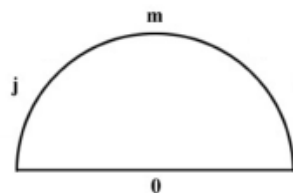
2.2.1.3 Complex Signs

Some signs can be decomposed into smaller pieces. We call them complex signs. These include agreement verbs, spatial verbs and classifier predicates.

2.2.1.3.1 Agreement Verbs

Agreement verbs are one subtype of lexical verbs in HKSL. This group of verbs can be marked for both subject-verb agreement and verb-object agreement via directing a sign towards different spatial loci in the signing space. The following figure shows the spatial arrangement in HKSL:

Figure 1 Spatial arrangement in HKSL



¹ The symbol [=] is from the CHAT format which means explanation. We indicate the description of the index signs with this symbol. This will allow researchers to have a better idea on the context of the signing.

Following Lam (2003), the alphabets *i*, *j*, *m* and *o* represent four spatial loci in the signing space (which is in front of the signer's torso). The alphabets *o* and *m* refer to the first and second person respectively. The alphabets *i* and *j* both indicate the third person in HKSL.

Agreement verbs can appear in three forms: (i) unmarked, (ii) directed to the actual location of the real referent or imagined referent and (iii) marked for verb agreement. These three forms are marked differently in the corpus.

(i) Unmarked form

When agreement verbs are not marked for verb agreement, the sign is glossed with its closest English translation: GIVE.

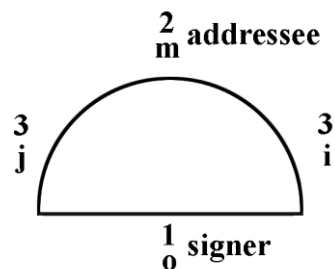
(ii) Directed to actual locations

If the agreement verbs are directed to actual/imagined locations, use a, b, c... to represent the locations.
e.g. GIVE-a

(iii) Marked for verb agreement

In Lam (2003), agreement values (i.e. first person, second person, etc.) are indicated by subscripts 1, 2 and 3. The actual location(s) observed in the agreement verbs are indicated by subscripts o, i, m and j. See the following figure:

Figure 2 Agreement markings in HKSL



Instead of using subscripts, hyphen is used to indicate the agreement values in gloss 1 and 2 tiers in the corpus (e.g. 1o-GIVE-3i).

The conventional representation of agreement in HKSL is used in the gloss 1 and 2 tiers in order to facilitate detailed study of verb agreement. A different system, however, is used in the main line in order to facilitate calculations in the CLAN program. Following the CHAT format, a fully marked agreement verb like 1o-GIVE-2m is represented as GIVE-1S&Sub-2S&Obj where the codes *Sub* and *Obj* denote subject-verb agreement and verb-object agreement, the person/number values (1, 2, 3/S,P) associated with the subject-verb agreement and the verb-object agreement is linked up with the symbol &.

Note that the spatial arrangements in Figure 2 may shift slight to the right or to the left when role shift (where the signer assumes the role of another person) takes place. Agreement is marked according to the shifted spatial arrangements in these cases, e.g.:

*SCH: IX_1 [% rs] GIVE-1S&Sub-3S&Obj BRENDA

2.2.1.3.2 Spatial Verbs

Spatial verbs denote the location of the referents. In order to distinguish person agreement from spatial location, we use a, b, c... to represent the location for spatial verbs, for instance, PUT-a PUT-b.

2.2.1.3.3 Classifier Predicates

Instead of decomposing the classifier predicates in the main line, glosses 1 and 2, we choose to gloss the classifier predicates based on meaning which is indicated with the symbol [=]. Only verb root is extracted at the beginning in the main gloss.

Three types of classifier predicates are also glossed in the main line, glosses 1 and 2:

Types	Codes	Examples
Semantic Classifiers	CL_sem	run+CL_sem:person [= a person runs]
Handle Classifiers	CL_hand	hit+CL_hand:sword [= hit with a sword]
SASS	CL_sass	round+CL_sass:xxx [= a round object]
Bodypart Classifiers / handshape representing bodypart	CL_body	walk+CL_body:person [= a person walks]

2.2.1.4 Numbers and Alphabets

Numbers are glossed as ONE, TWO, etc. Sign alphabets are glossed as A, B, C.

2.2.1.5 Listing

It is common for HKSL signers to express ‘first, second, third, etc.’ or ‘number one, number two, number three’ as one single sign. We call these signs listing signs and they are glossed as:

FOUR_LIST+IX_ONE FOUR_LIST+IX_TWO FOUR_LIST+IX_THREE FOUR_LIST+IX_FOUR
‘there are four things, number one, number two, number three and number four’

or

‘there are four things, the first one, the second one, the third one and the fourth one’

2.2.1.6 Compounds

Compounds are glossed as one single sign in the main line.

2.2.1.7 Contracted Signs

It is common for signers to contract two signs into one sign. They are like contracted form *isn't* in English. These signs are glossed as DO~WHAT ‘do what’, LIKE~DISLIKE ‘like or dislike’.

2.2.1.8 Gesture

It is common for gesture to interweave with signs in adult signing. They may also be early forms of signs in child language. Therefore, unlike the spoken language corpus, we include gesture in the transcription where necessary.² Akin to the transcription for classifier predicates, index signs, we describe the meaning of the gesture with the symbol [=].

Gestures are glossed as: gesture [= play toy].

² The symbol [+ bch] is added to allow the researchers to include or exclude the utterances in the calculations for mlt and mlv. See the CHAT Transcription Manual for further details.

2.2.1.9 Mouthing

Due to the strong oralist atmosphere in Hong Kong, it is common for signers to mouth proper names or new terms. Meaningful mouthing glossed as mouthing: *meaning of the mouthing* (e.g. as mouthing:Leung_Ka_Fu).

2.2.2 Special coding

In addition to the basic transcription, we also use some markers to denote the special forms. The following sections list the markers used in the Child HKSL Corpus.

2.2.2.1 Adult signs

2.2.2.1.1 Family specific forms

Some signs are used within the child's family, but not in the deaf community. These signs are glossed with a marker @f. e.g. CANDY@f

2.2.2.1.2 Child-directed forms

Adults may modify their signs when they sign to a child, these signs are called child-directed signs. The following are some cases where adult signs are glossed as child-directed signs, marked with @m.

Cases	Examples
An adult manipulates a child's hand to sign the adult form	NOODLE@m
An adult signs a sign at the child's body	DUCK@m
An adult demonstrates the handshape of a sign to a child	RICE@m
An adult follows the family specific forms or child-invented forms	CANDY@m
An adult imitates the signs used by the child	SNEEZE@m

2.2.2.2 Child signs

2.2.2.2.1 Child-invented forms

Children may invent signs which look different from the adult forms. Gloss them with a special form marker @c. (e.g. HARD@c 'hard')

2.2.2.2.2 Substitution

Children substitute a sign for another. For instance, a child intends to sign GREEN 'green', but he produces BLUE 'blue'. Then gloss the sign as: BLUE [* GREEN] 'blue, but the child's intended meaning is green'.

2.2.2.2.3 Imitation³

It is common for children to imitate the adult signs. We gloss these signs with [“] in the gloss 1 and gloss 2 tiers to mark imitation, e.g. CANDY [“] 'candy'. The postcode [+ imit] is added in the main line if all the words in the utterance are imitations. This allows researchers to choose between including utterances with partial imitation and utterances with full imitation.

2.2.2.3 Others

2.2.2.3.1 Repetition

Repeated signs are glossed with [x number of times], e.g. EGG [x 5]. 'egg'.

³ See footnote 2.

2.2.2.3.2 Unintelligible Signs

Unintelligible materials are glossed with xxx:

Sign:	XXX
Gesture:	gesture [= xxx]
Index sign:	IX [= xxx]

2.2.2.3.3 Signs which are out of the screen

Participants may be out of the screen occasionally. Part of the signs may be seen. Some of the signs are totally out of sight. These signs are glossed as WWW. The best guess of the signs would be noted in a comment bracket, e.g. WWW [% WHAT].

2.2.2.3.4 Incomplete Signs

Incomplete signs are glossed with the symbol &, for example: &CANDY ‘candy’.

2.2.2.3.5 Incorrect forms

The symbol [*] marks the phonetic error of a sign. This may occur in both the child tier and the hearing researcher tier, e.g. CANDY [*] ‘candy’.

2.2.2.3.6 Alternative forms

Some signs have alternative transcription and we gloss them as: e.g. EAT [= ? FOOD] ‘eat or food’.

2.2.3 Comments on signs

2.2.3.1 Idiosyncratic forms

Some adult forms may be signs from the idiolect. We gloss these signs like the following:

NOT [% idiosyncratic]

2.2.3.2 Fast signing and assimilation

Phonetic forms of a sign may deviate from those in conventional HKSL. These signs are glossed with a comment in the square bracket, e.g. TELL [% fast signing], CAKE [% assimilation with SIGN].

2.2.3.3 Handedness

When a one-handed sign becomes a two-handed sign, a comment is added by using the % symbol in brackets: e.g. EAT [% two-handed].

Sometimes two-handed signs are uttered with just one hand as the signer may use one of his/her hands to hold some objects (e.g. book). These signs are also followed by a comment in the square bracket, e.g. CAT [% hold book].

2.2.3.4 Role Shift

Role shift is glossed as “rs” in the comment bracket, e.g. [% rs]

2.2.3.5 Slip of the hand

When a deaf adult produces a wrong sign because of the slip of the hand, we gloss it as CANDY [% slip of the hand].

2.2.4 Utterance properties glossed in the Main Line

2.2.4.1 Incomplete utterances

Following the CHAT manual, utterances are delimited with . ? !. The symbol +/- marks incomplete utterances.

2.2.4.2 Repeated phrases

Phrases may be repeated in actual signing. This phenomenon is glossed with the symbols <SIGN SIGN SIGN> [/], e.g.:

<IX_obj [= picture] WHAT IX_obj [= picture] WHAT> [/] IX_obj [= picture] WHAT?

2.2.4.3 Simultaneity

Signs may be articulated using two hands. It is possible for a signer to sign two separate signs with the two hands at the same time. There are several kinds of simultaneity in HKSL.

(i) Unrelated signs

When the two signs uttered simultaneously are not related, we link them using the a ~ and the angle brackets followed by a comment < > [% sim] show that the ordering in the main line is not relevant as the two signs are simultaneously produced.

*CHI: <HAVE~WITCH> [% sim]
'There is a witch.'

(ii) Related signs

Sometimes the simultaneous signs are considered as one single sign. Then these signs are linked up with the symbol + in the main line, e.g.:

*KEN: IX_2+SELF
'yourself / you alone'

Sometimes a sign A continues to hold in the signing space to combine with another sign articulated by another hand to form a polymorphemic sign. The symbol &{|=SIGN marked the beginning time of holding while &|=SIGN marked the end of holding.

be_located+CL_sem:thief [= a thief] &{|=CL_sem:thief POLICE
fire_a_gun+CL_sem:gun [= the police fire a gun to the person] &|=CL_sem:thief

The semantic classifier CL_sem is held when the signer continue to sign POLICE and fire_a_gun+CL_sem:gun .

(iii) Discourse Buoy

We use > to represent a sign which continues to hold in the signing space. These annotations would not appear in the main line. This symbol only appears in the tiers *gloss 1 and gloss 2*. When a sign continues to hold in the following utterance, it is marked with >>. If the sign continues to hold in the next following utterance, it is marked with >>>. The number of the symbol > marks the number of the utterances that involve discourse buoy.

2.3 Tagging

2.3.1 Grammatical Categories

Category	Code	Example
Adjective	adj	SHAMEFUL, SHINY, SHORT
Adverb	adv	NOW, PLEASE, SAME
Agreement verb	v:agr	GIVE, TELL, HELP
Classifier predicates	cl	be_located+CL_sem:thief [= a thief]
Conjunction	conj	BUT, IF
Interjection	inter	GOOD, CORRECT
Modal	mod	CAN, WON'T
Negation marker	neg	NOT, NOT_HAVE
Noun	n	TELEPHONE, BALL, DOG
Number	num	ONE, TWO
Personal pronoun	n:pro	IX_1, IX_2
Plain verb	v:pl	KNOW, EAT
Preposition	prep	ABOVE, BELOW
Proper name	n:pr	KENNY, BRENDA
Quantifier	quant	MANY, ALL
Question word	que	GOOD_NOT_GOOD, YES_NO_YES
Spatial verb	v:sp	PUT, WALK
Wh-word	wh	WHAT, WHY, HOW, WHO

2.3.2 Functional Tagging

Category	Code	Example
Gesture	ges	ges [= play toy]
Fingerspelled letters	alp	A, B, C
Index sign	ix:obj	IX_obj [= picture]
	ix:loc	IX_loc [= kitchen]
	ix:cmd	IX_cmd [= go there]
	ix:evt	IX_evt [= younger sister cries]
mouthing in Cantonese	ca	mouthing:leung_ka_fu

2.3.3 Complex Tagging

2.3.3.1 Contracted Forms

Categories of each sign in the contracted form would be tagged independently and linked up with ~, e.g. v:pl|DO~wh|WHAT.

2.3.3.2 Simultaneous Signs

Unrelated simultaneous signs are tagged in the same way as the contracted form by using the symbol ~, e.g. mod|HAVE~n|WITCH.

Related simultaneous signs, which together form a larger morphosyntactic constituent are tagged in the following form: n:pro|+n:pro|IX_2+n:pro|SELF.

2.3.4 Alternative Tagging

A sign may belong to two or more categories in the same context, we provide alternative tagging by using the symbol ^, e.g.: v:pl|SIGN^n|SIGN.

Draft

2.3.5 Morphological Tagging

2.3.5.1 Verb Agreement

Verb agreement is v:agr|GIVE-1S&Sub-2S&Obj

CENTRE FOR SIGN LINGUISTICS & DEAF STUDIES
CHILD HKSL CORPUS MANUAL - Draft

Appendix 1 – Symbol Summary

A. Symbols in main line

Symbols	Meaning	Gloss in Main Line	Tagging in % mor tier	Sections
IX	Index signs	IX_3	n:pro IX_3	2.2.1.2
_1 _2 _3 _1p _2p _3p	Agreement markings (first, second and third; ‘p’ stands for plural)	IX_cmd [= sit properly]	ix:cmd IX_cmd	2.2.1.2 2.3.2
-a -b -c	Location markings	PUT-a	v:sp PUT	2.2.1.3.2
S &Sub &Obj	Person verb agreement (‘S’ stands for singular)	GIVE-2S&Sub-1S&Obj	v:agr GIVE-2S&Sub-1S&Obj	2.2.1.3.1 2.3.5.1
CL	Classifier predicates	move+CL_sem [= a car moves forward]	cl +v move+cl CL_sem	2.2.1.3.3
~	Contracted form markings	DO~WHAT	v:pl DO~wh WHAT	2.2.1.7 2.3.3.1
	Unrelated simultaneous signs	<IX_obj~HAIR [= the hair]> [% sim]	ix:obj IX_obj~n:HAIR	2.2.4.3 2.3.3.2
gesture	Gestures which may be early forms of signs	gesture [= play toy]	ges gesture	2.2.1.8
mouthing	Mouthing	mouthing:leung_ka_fu	ca mouthing	2.2.1.9
@f	Family-specific form markings	CANDY@f	n CANDY@f	2.2.2.1.1
@m	Child-directed form markings	DUCK@m	n duck@m	2.2.2.1.2
@c	Child-invented form markings	HARD@c	adj HARD@c	2.2.2.2.1
[* SIGN]	Substitution of signs	BLUE [* GREEN]	adj BLUE	2.2.2.2.2
[_imit]	Child-imitated form markings	CANDY [+ imit] .	NR	2.2.2.2.3
[x N]	Repetition of the same signs	EGG [x 5]	n EGG	2.2.2.3.1
XXX	Unintelligible signs / gestures / index	XXX gesture [= xxx] IX_obj [= xxx]	unk XXX ges gesture ix:obj IX_obj	2.2.2.3.2
WWW	Signs which are out of the screen	WWW [% WHAT]	unk www	2.2.2.3.3
&	Incomplete signs	&CANDY	NR	2.2.2.3.4
[*]	Incorrect form markings	CANDY [*]	n CANDY	2.2.2.3.5
[=?]	Alternative transcription	EAT [=? FOOD]	v EAT	2.2.2.3.6
[%]	Comments	NOT [% idiosyncractic]	neg NOT	2.2.3
+/.	Incomplete utterance markings	YOUNGER_SISTER SIT +/.	n YOUNGER_SISTER v:pl SIT +/.	2.2.4.1
<> [/]	Repeated phrases	<IX_obj [= picture] WHAT> [/] IX_obj [= picture] WHAT ?	ix:obj PICTURE wh WHAT ?	2.2.4.2
< ~ > [% sim]	Unrelated simultaneous signs	<IX_2~SELF> [% sim]	n:pro IX_2~n SELF	2.2.4.3
&{ =x &} =x	Continuous form markings of polymorphemic signs	be_located+CL_sem [= a person] &{ =CL_sem POLICE FIRE_A_GUN- a+CL_sem [= the police fire a fun to the person] &} =CL_sem	cl +v be_located+cl CL_se m n POLICE cl +v FIRE_A_GUN+cl CL _sem	2.2.4.3

Draft

B. Symbols appear only in the tiers *gloss 1* and *gloss 2*

Symbols	Meaning	Gloss	Tagging in %mor tier	Sections
-1o, -2m, -3i, -3j	Agreement markings	1o-GIVE-3i	NR	2.2.1.3.1
[“]	Child-imitated form markings	CANDY [“]	NR	2.2.2.2.3
>, >>, >>>, etc.	Discourse buoy	NR	NR	2.2.4.3

C. Symbols appear in %mor tier

Symbols	Meaning	Gloss	Tagging in %mor tier	Sections
^	Alternative tagging	CUT_HAIR	v:pl CUT_HAIR^v:sp CUT_HAIR	2.3.4

Notes: Blue words = symbols created by the research group

References

Lam, W. S. 2003. Verb Agreement in Hong Kong Sign Language. MPhil thesis. The Chinese University of Hong Kong.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates