RADIS Corpus of Spanish Sign Language (LSE)

Glosses Annotation Conventions

Version 2.1 – May 2018 (revised January 2024)

Ania Pérez, José Mª García-Miguel and Carmen Cabeza

Universidade de Vigo

. Contenido

MD_Glosa and MI_Glosa tiers

MD_Glosa (right-hand gloss) and MI_Glosa (left-hand gloss) tiers contain only the Spanish annotation of the manual component, which can be lexical (1.1) or semi-lexical/non-lexical (1.2). RH_Gloss and LH_Gloss contain the corresponding glosses in English, with some simplification.

Lexical signs are dictionary signs. They are annotated with ID-glosses. If they are bimanual, the same ID-gloss is used on both lines. Here an English version is provided:

> (1) MD_Glosa: NAME
>   MI_Glosa: NAME

Semi-lexical/non-lexical signs include indexes, buoys, classifiers, gestures, and fingerspelling. If each hand produces a separate meaningful unit, the annotation of the lines is different:

> (2) MD_Glosa: cl:PERSON-sit-floor
>   MI_Glosa: cl:SOIL

1. ID-glosses of lexical signs

Lexical signs are annotated by means of ID-glosses (**identifying glosses**). An ID-gloss is a label word that systematically refers to a sign, even if it has an inflectional variation (cf. figs. 1a and 1b). In addition, the ID-glosses link the forms of the corpus to the lexical database[1].

(a)	(b)
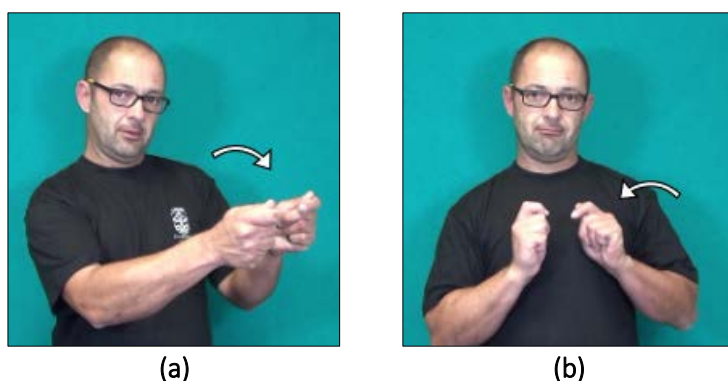
Fig. 1. Inflectional variants of ID-gloss GIVE(2H).

However, there are three types of variation that do result in the creation of more than one ID-gloss of the same sign: phonetic variation, numeral incorporation and the involvement of hands (1.1.1). In these cases, the systematic reference to the sign is safeguarded through the **lemma**, which is the core of the ID-gloss. Table 1 shows the structure of the ID-gloss and the spelling conventions at work in its creation.

---

[1] The lexical database linked to the corpus can be accessed on the http://isignos.uvigo.gal/lexico / website.

Table 1. ID-gloss: structure and annotation conventions

ID-gloss structure:

Lemma	Variation 1 Phonetics	Variation 2 Numeral incorporation	Variation 3 Hand involvement	→ ID-glosa	→ ID-gloss
DAR	(L)		(2M)	DAR(L)(2M)	GIVE(L)(2H)
MES		(DOS)	(MP)	MES(DOS)(MP)	MONTH(TWO)(PH)

Element	Meaning	Ejemplo	Example
-	Separator - lemmas with more than one word	HABER-NO A-POR-TODAS	HAVE-NOT GO-FOR-IT
^	Separator ^ compound signs	PADRE^MADRE	FATHER^MOTHER
.	Separator . "hint", allows to use two words in the creation of a lemma to facilitate the identification of the sign	MÁS.SUMA MÁS.AÑADIR MÁS.CANTIDAD	MORE.SUM MORE.ADD MORE.AMOUNT
Numbers	They allow you to create different lemmas with the same word	PERA2 PERA3	PEAR2 PEAR3

1.1. Phonetic variation, numeral incorporation and hand involvement

There are three types of variation that result in the creation of more than one ID-gloss of the same sign: phonetic variation, numeral incorporation, and hand involvement.

Phonetic variation. At the phonetic level, we consider that two signs are variants when, in addition to sharing the meaning, they differ in an articulatory parameter. Phonetic variation is annotated with an alphanumeric code in parentheses. For example: FATHER and FATHER(2p) have the same meaning and are articulated in the same way except for the handshape, which is indicated in parentheses "(2p)".

PADRE   PADRE(2p)   2p =
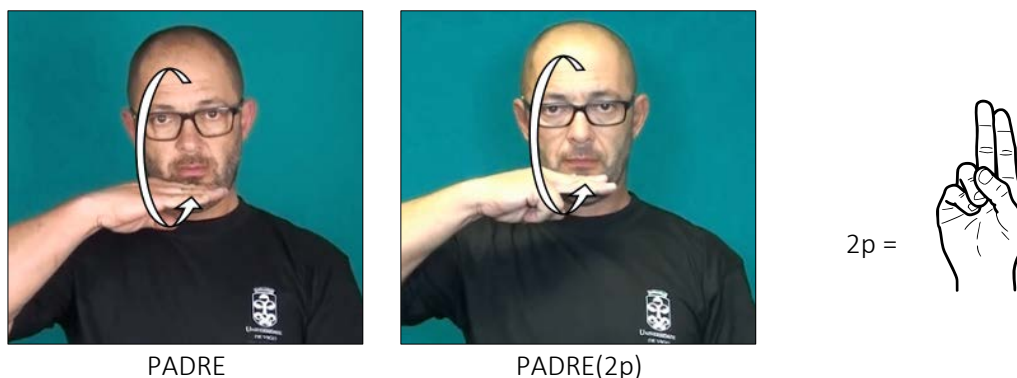
Fig. 2. FATHER and FATHER(2p) ID-glosses

Numeral incorporation. If a number is incorporated into a sign, it is indicated by capital letters in parentheses. If the citation form corresponds to the numeral ONE, it is not added, for example YEAR.FUTURE, YEAR.FUTURE(TWO).

AÑO.FUTURO   AÑO.FUTURO(DOS)   DOS =

Fig. 3. YEAR.FUTURE and YEAR.FUTURE(TWO) ID-glosses

Hand involvement. By hand involvement we mean variation in the number of manual articulators and in the relationship of dominance between them reflected in the movement. It is annotated by the following code:

1M	1H	One hand
2M	2H	Two hands with simoultaneous movement
2Msu	2Hsuc	Two hands with succesive movement
MP	PH	Two hands with passive hand

1.2. Negation

Some signs incorporate negation without the need to use the NO sign. These signs have their own entry in the lexical database and are glossed with the suffix "NO" as follows: HABER-NO, NECESITAR-NO (HAVE-NOT, NEED-NOT).

1.3. Composition

Composition is the process of forming new signs through the union of two or more lexemes. Their meaning is not predictable from the signs that make them up. The compounds are glossed with the same ID-glosses of the signs that form them, separated by "^", giving rise to a new ID-gloss, for example:: MENTE^NORMAL (MIND^NORMAL, 'innocent, ignorant').

1.4. Numbers

Numbers are glossed with words, for example: UNO, PRIMERO (ONE, FIRST). If a number is incorporated into a sign, it is indicated by capital letters in parentheses. (see section 1.1.1 *Phonetic variation, numeral incorporation and hand involvement*).

1.5. Signs of other SL

Signs from other sign languages are also glossed and included in the lexical database, in which the language of origin is indicated. To date, the only language with these characteristics is LAG (Galician Sign Language).

1.6. Interrupted and non-decipherable signs

Some signs of the corpus are interrupted in their execution without being completed. If the interruption is an inchoative aspect mark, the ID-gloss is noted normally. Otherwise, a hyphen "-" is added to the ID-gloss, like this: APPLE-

In the case of indecipherable joints, the label "hhh" is used.

The iSignos web offers a list of LSE lexicon alphabetically ordered, in Spanish and English: http://isignos.uvigo.es/es/lexico.

2. Annotation of semi-lexical/non-lexical signs

2.1. Basic pattern of annotation

Indexes, buoys, classifiers and gestures are annotated following the same pattern in order to simplify and facilitate the annotation process:

category.type(handshape):meaning

For example: INDX.PRO(B):1sg is the annotation of an index (category) pronominal (type) of B handshape that refers to a first person singular (meaning).

The pattern marks the order of the elements and how they should be separated graphically, but does not prescribe the annotation of all of them.

Tabla 2. Basic annotation pattern

category	.	type	(handshape)	:	meaning
INDX	.	PRO	(B)	:	1sg
cl	.	D	(1)	:	Window-shape

Elemento	Significado
.	Separator . to indicate the type within each category
( )	Sign handshape is indicated in parentheses
:	Separator : precedes the meaning of the sign
-	Word breaker (in meaning)

In tiers RH_Gloss y LH_Gloss just the presence of classifiers is annotated with *cl* abreviation.

2.2. Indexes

All indexes are annotated with the label "INDX". The types of indexes annotated are as follows:

5

Todos los índices son anotados con la etiqueta "INDX". Los tipos de índices anotados son los siguientes:

INDX.PRO   (pronominal index)
INDX.DET   (determiner index)
INDX.LOC   (locative index)
INDX.AUX[2]   (auxiliar index)

2.3. Buoys

Buoys are one-hand handshapes that remain static while the other continues to sign. They are annotated with the abbreviation "B". As a meaning, the ID-gloss of the sign that acts as a buoy is noted, for example: B:BICYCLE, B:INDX.

If the buoy is made up of a number of outstretched fingers that correspond to ideas or entities, we are dealing with a **list buoy**. The type of buoy is indicated by the abbreviation "L". The dominant hand is glossed as an index and the element of the list buoy referred to is included in its meaning. In (3) we show the annotation of 'the second of three elements':

Si la boya está formada por un número de dedos estirados que se corresponden con ideas o entidades estamos ante una **boya lista**. Se indica el tipo de boya con la abreviatura "L". La mano dominante es glosada como índice y se recoge en su significado el elemento de la boya lista referido. En (3) mostramos la anotación de 'el segundo de tres elementos':

(1)   MD_Glosa: INDX:segundo   RH_Gloss: INDX: second
       MI_Glosa: B.L(3)   LH_Gloss: L.B(3)

2.4. Classifier constructions

Classificatory predicates are annotated with the label "cl". The types of classifier constructions are annotated as follows:

cl.e   (entity)
cl.m   (manipulative)
cl.d   (descriptive)
cl.c   (body)

The **annotation of meaning** in classifier predicates is done schematically. Articles, nexus and prepositions are dispensed with. The use of verb forms is reduced to infinitive and participle. Discourse referents are also *generalized*, e.g. "cl.e(3):persons3" and not "cl.e(3):one-boy-and-two-girls".

If the classifier for each hand is part of a set, the meaning should reflect the set, while the handshape indicates the manual distribution of the entities:

(2)   MD_Glosa: cl.e(2):PERSONAS7-caminar   cl.e(2):PERSONS7-walk
       MI_Glosa: cl.e(5):PERSONAS7-caminar   cl.e(5):PERSONS7-walk

---

[2] Directional auxiliary, it implies two grammatical loci.

The annotation of meaning has up to three constituents. In general, we can say that the first indicates the meaning assigned to the manual handshape, the second the predication, and the third the referent of the dominated hand.

(3)   MD_Glosa: cl.d(5):agua-llenar-olla   cl.d(5):water-fill-pot
        MI_Glosa: cl.d(Xc):agua-llenar-olla   cl.d(Xc): wáter-fill-pot

These conventions just apply to the ID-glosses in Spanish. In tiers RH_Gloss and LH_Gloss the presence of classifiers is annotated with the abbreviation *cl*, without any more information.

2.5. Gestures

Gestures are annotated with the label "G" and a brief description of the meaning of the gesture in lowercase letters, like this: G:no-entender (G:no-understand)

2.6. Fingerspelling

Spellings through the manual alphabet are annotated with the prefix DL, followed by ":" and the spelled word in capital letters, like this: DL:NICARAGUA.

The fingerspelling annotation corresponds to what is articulated by the signer, regardless of whether letters are omitted or there is a lapse.

Some signs, for whatever reason, proceed from and coincide with letters or sequences of letters in the dactylology. They are included in the lexical database and annotated with their respective ID-glosses. For example: Y, O, FP[3] ('and', 'or', 'VET').

2.7. Proper nouns

Name signs are entered with the abbreviation SN (*NS* in English gloss lines), followed by the separator ":" and the proper noun, e.g. *SN*:KUMA.

The signs of institutionalized names (places, organizations and relevant public figures in the Galicial Deaf community) are glosses and incorporated into the lexical database, even if they coincide in forma with another lexical sign, for example: A-CORUÑA, ACUÑA (the name of a school and a famous Galician Deaf sculptor).

---

[3] Formación Profesional (FP): Vocational Education and Training (VET).