

Convenciones de Anotación

Corpus RADIS de la LSE

Versión 2 – mayo de 2018



Ania Pérez, José M^a García-Miguel y Carmen Cabeza

Universidade de Vigo

1. INTRODUCCIÓN	2
2. LÍNEAS D E I	2
1.1. ID-GLOSAS DE SIGNOS LÉXICOS	2
1.1.1. <i>Variación fonética, incorporación numeral y manualidad</i>	3
1.1.2. <i>Negación</i>	4
1.1.3. <i>Composición</i>	4
1.1.4. <i>Números</i>	4
1.1.5. <i>Signos de otras LS</i>	5
1.1.6. <i>Signos interrumpidos y no descifrables</i>	5
1.2. ANOTACIÓN DE SIGNOS SEMILÉXICOS/NO LÉXICOS	5
1.2.1. <i>Patrón básico de anotación</i>	5
1.2.2. <i>Índices</i>	5
1.2.3. <i>Boyas</i>	6
1.2.4. <i>Construcciones con clasificadores</i>	6
1.2.5. <i>Gestos</i>	7
1.2.6. <i>Dactilológico</i>	7
1.2.7. <i>Nombres propios</i>	7
3. LÍNEA T	8

1. Introducción

Este documento describe las convenciones en la anotación del Corpus RADIS de la LSE.

La compilación y anotación de este corpus se inició en el marco del proyecto de investigación RADIS (Relaciones Actanciales en Discurso Signado). Su objeto era el estudio de los procedimientos gramaticales que utiliza la LSE para marcar las relaciones semánticas entre predicados y participantes, con base en una metodología de corpus.

2. Líneas D e I

En las líneas D (mano derecha) e I (mano izquierda) figura exclusivamente la anotación del componente manual, que puede ser léxico (1.1) o semiléxico/no léxico (1.2).

Los **signos léxicos** son signos de diccionario. Se anotan con ID-glosas. Si son bimanuales se usa la misma ID-glosa en ambas líneas:

- (1) D: NOMBRE
I: NOMBRE

Los **signos semiléxicos/no léxicos** incluyen índices, boyas, clasificadores, gestos y dactilología. Si cada mano produce una unidad significativa separada la anotación de las líneas es diferente:

- (2) D: cl:PERSONA-sentarse-suelo
I: cl:SUELO

1.1. ID-glosas de signos léxicos

Los signos léxicos son anotados a través de ID-glosas (**glosas identificativas**). Una ID-glosa es una palabra etiqueta que remite sistemáticamente a un signo, aunque éste presente variación flexiva o de número (cf. figs. 1a y 1b)¹. Además las ID-glosas vinculan las formas del corpus con la base de datos léxica².

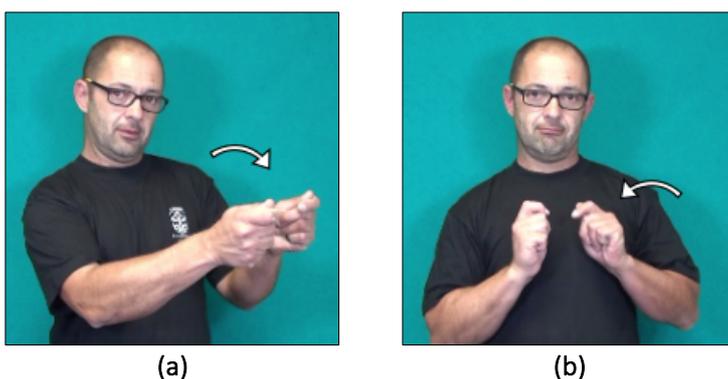


Fig. 1. Variantes flexivas de ID-glosa DAR(2M). (a) 1P = agente. (b) 1P = receptor

¹ Esta variación morfológica (flexión y número) no se refleja en la ID-glosa, pero sí se registra en otras líneas. la fase de anotación secundaria en líneas específicas.

² En la web <http://nasgrades.uvigo.es/LSE-web/lexico/lex.php> se puede acceder a la base de datos léxica ligada al corpus.

Sin embargo hay tres tipos de variación que sí producen la creación de más de una ID-glosa del mismo signo: variación fonética, incorporación numeral y manualidad (1.1.1). En estos casos se salvaguarda la referencia sistemática al signo a través del **lema**, que es el núcleo de la ID-glosa. La tabla 1 muestra la estructura de la ID-glosa y las convenciones ortotipográficas operativas en su creación.

Tabla 1. ID-glosa: estructura y convenciones ortotipográficas

Estructura ID-glosa:

Lema	Variación 1 fonética	Variación 2 incorporación numeral	Variación 3 manualidad	→ ID-glosa
DAR	(L)		(2M)	→ DAR(L)(2M)
MES		(DOS)	(MP)	→ MES(DOS)(MP)

Elemento	Significado	Ejemplo
-	Separador: lemas con más de una palabra	HABER-NO A-POR-TODAS
^	Separador: signos compuestos	PADRE^MADRE
.	Separador: “pista”, permite usar dos palabras en la creación de un lema para facilitar la identificación del signo	MÁS.SUMA MÁS.AÑADIR MÁS.CANTIDAD
numerales	Permiten crear lemas diferentes con la misma palabra	PERA2 PERA3

1.1.1. Variación fonética, incorporación numeral y manualidad

Hay tres tipos de variación que producen la creación de más de una ID-glosa del mismo signo: variación fonética, incorporación numeral y manualidad.

Variación fonética. En el plano fonético consideramos que dos signos son variantes cuando, además de compartir el significado, difieren en un parámetro articulatorio. La variación fonética se anota con un código alfanumérico entre paréntesis. Por ejemplo: PADRE y PADRE(2p) tienen el mismo significado y se articulan igual a excepción de la configuración, que se indica entre paréntesis “(2p)”.



Fig. 2. ID-glosas PADRE y PADRE(2p)

Incorporación numeral. Si un número es incorporado a un signo se indica con letras mayúsculas entre paréntesis. Si la forma de cita se corresponde con el numeral UNO no se añade, por ejemplo [AÑO.FUTURO](#), [AÑO.FUTURO\(DOS\)](#).



Fig. 3. ID-glosas [AÑO.FUTURO](#) y [AÑO.FUTURO\(DOS\)](#)

Manualidad. Por manualidad entendemos variación en el número de articuladores manuales y en la relación de dominancia entre ellos reflejada en el movimiento. Se anota mediante el siguiente código:

- 1M (monomanual)
- 2M (bimanual simultáneo)
- 2Msu (bimanual sucesivo)
- MP (bimanual con mano pasiva)

1.1.2. Negación

Algunos signos incorporan la negación sin necesidad de emplear el signo [NO](#). Estos signos tienen su propia entrada en la base de datos léxica y son glosados con el sufijo “NO” de la siguiente forma: [HABER-NO](#), [HACER-FALTA-NO](#).

1.1.3. Composición

La composición es el proceso de formación de nuevos signos a través de la unión de dos o más lexemas. Su significado no es predecible de los signos que los conforman. Los compuestos son glosados con las mismas ID-glosas de los signos que los forman, separadas por “^”, dando lugar a una nueva ID-glosa, por ejemplo: [MENTE^NORMAL](#) (‘inocente, ignorante’).

1.1.4. Números

Los números son glosados con palabras, no con números, por ejemplo: [UNO](#), [PRIMERO](#). Si un número es incorporado a un signo se indica con letras mayúsculas entre paréntesis. (v. apartado 1.1.1 *Variación fonética, incorporación numeral y manualidad*).

1.1.5. Signos de otras LS

Los signos de otras lenguas de signos son glosados e incluidos igualmente en la base de datos léxica, en la que se indica la lengua de procedencia. Hasta la fecha la única lengua de estas características es la LAG (Lingua de Acenos Galega).

1.1.6. Signos interrumpidos y no descifrables

Algunos signos del corpus son interrumpidos en su ejecución sin llegar a completarse. Si la interrupción es una marca de aspecto incoativo es anotada la ID-glosa con normalidad. En caso contrario se añade un guion “-” a la ID-glosa, así: MANZANA-

En el caso de articulaciones no descifrables se emplea la etiqueta “hhh”.

1.2. Anotación de signos semiléxicos/no léxicos

1.2.1. Patrón básico de anotación

Índices, boyas, clasificadores y gestos son anotados siguiendo el mismo patrón con el objetivo de simplificar y facilitar el proceso de anotación:

categoría.tipo(configuración):significado

Por ejemplo: INDX.PRO(B):1sg es la anotación de un índice (categoría) pronominal (tipo) de configuración B que remite a una primera persona de singular (significado).

El patrón marca el orden de los elementos y cómo deben ser separados gráficamente, pero no prescribe la anotación de todos ellos.

Las **configuraciones manuales** se anotan mediante el código que se incluye en el Apéndice II.

Tabla 2. Patrón básico de anotación

categoría . tipo (configuración) : significado	
INDX . PRO (B) : 1sg	
cl . d (1) : forma-ventana	
Elemento	Significado
.	Separador: para indicar el tipo dentro de cada categoría
()	Entre paréntesis se indica la configuración manual del signo
:	Separador: precede el significado del signo
-	Separador de palabras (en el significado)

1.2.2. Índices

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

- INDX.PRO (índice pronombre)
- INDX.DET (índice determinante)
- INDX.LOC (índice locativo)

INDX.AUX³ (índice auxiliar)

1.2.3. Boyas

Las boyas son configuraciones de una mano que permanecen estáticas mientras la otra sigue signando. Son anotadas con la abreviatura “B”. Como significado se anota la ID-glosa del signo que actúa como boya, por ejemplo: B:BICICLETA, B:INDX.

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’:

(3) D: INDX:segundo
I: B.L(3)

1.2.4. Construcciones con clasificadores

Los predicados clasificatorios son anotados con la etiqueta “cl”. Los tipos de construcciones con clasificador anotados son los siguientes (v. Apéndice III):

cl.e (entidad)
cl.m (manipulativo)
cl.d (descriptivo)
cl.c (corporal)

La **anotación del significado** en predicados clasificatorios se hace de forma esquemática. Se prescinde de artículos, nexos y preposiciones. El uso de formas verbales se reduce a infinitivo y participio. Los referentes discursivos sufren también una *generalización*, por ejemplo “cl.e(3):personas3” y no “cl.e(3):un-niño-y-dos-niñas”.

Si el clasificador de cada mano es la parte de un conjunto el significado debe reflejar el conjunto, en tanto que la configuración indica la distribución manual de las entidades:

(4) D: cl.e(2):PERSONAS7-caminar
I: cl.e(5):PERSONAS7-caminar

La anotación del significado cuenta con hasta tres constituyentes. En general podemos decir que el primero indica el significado asignado a la configuración manual, el segundo la predicación y el tercero el referente de la mano dominada (v. Apéndice III).

(5) D: cl.d(5):agua-llenar-olla
I: cl.d(Xc):agua-llenar-olla

A pesar de estas convenciones generales existen **peculiaridades en la anotación del significado**. Pueden verse en el Apéndice III, un esquema con ejemplos del corpus que ilustra la anotación de cada tipo y subtipo de construcción con clasificador. En el caso del tipo entidad es necesario además hacer alguna aclaración:

³ Auxiliar direccional, implica dos lugares gramaticales.

- *Construcciones con clasificadores de entidad.* El número de constituyentes que se anota en cada mano refleja el tipo de relación entre ellas. En el Apéndice IV se trata con detalle esta cuestión. El primer constituyente (la entidad representada en el clasificador) se anota en letras mayúsculas. Las palabras empleadas para anotar las acciones proceden de un inventario reducido, aunque abierto (v. Apéndice V). Por ejemplo, en (6) la anotación nos indica que la construcción es de activa/pasiva y que la pasiva es estática.

(6) D: cl.e(2d):PERSONA-sentarse-suelo
 l: cl.e(Bh):SUELO

1.2.5. Gestos

Los gestos son anotados con la etiqueta “G” y una breve descripción del significado del gesto en letras minúsculas, así: G:no-entender. Los tipos de índices anotados son los siguientes:

G.EMB (gesto emblema)
 G.ILU (gesto ilustrador)
 G.EMO (gesto muestra de emoción)
 G.REG (gesto regulador)
 G.ADA (gesto adaptador)

1.2.6. Dactilológico

Los deletreos a través del alfabeto dactilológico manual son anotados con el prefijo DT, seguido de “:” y la palabra en letras mayúsculas, así: DT:ANA.

La anotación del dactilológico se corresponde con lo articulado por la persona signante, independientemente de que en el deletreo se omitan letras o exista algún lapsus.

En deletreos de más de una palabra la anotación se realiza en el mismo comentario, por ejemplo: DT:RUAESTRADA. Si las palabras se separan, bien mediante una pausa, bien a través de un signo separador (habitualmente mano en B, movimiento de desplazamiento) es anotado mediante el guion “-”, por ejemplo: DT:ANA-PZ.

Algunos signos, por el motivo que sea, proceden y coinciden con letras o secuencias de letras del dactilológico. Son incluidos en la base de datos léxica y anotados con sus respectivas ID-glosas. Por ejemplo: Y, O, FP.

1.2.7. Nombres propios

Los signos de nombres son introducidos con la abreviatura SN, seguida por el separador “:” y el nombre propio, por ejemplo SN:María. Esto es anotado así independientemente de que el signo de nombre de María sea idéntico en forma a un signo léxico.

Los signos de nombres institucionalizados (lugares, organizaciones y personajes públicos relevantes en la comunidad sorda gallega) son glosados e incorporados a la base de datos léxica, aunque coincidan en forma con otro signo léxico, por ejemplo: A-CORUÑA, ACUÑA (el nombre de un colegio).

3. Línea T

La línea T (traducción) proporciona una traducción literal lo más cercana posible a la estructura de la LSE. La intención es que todos los elementos **manuales** presentes en el texto signado tengan su correspondencia en la traducción. La información que **no procede de elementos manuales del** segmento traducido se anota entre paréntesis. Por ejemplo, en (7) se ha traducido entre paréntesis “pícaro”, que se expresa facialmente en LSE, y “al hombre”, que procede del lugar hacia el que se dirige el signo MIRAR.

(7) T: mira (pícaro al hombre),
D: MIRAR
I:

Otros elementos **proprios de la estructura de la lengua oral** que no son expresados manualmente en LSE también son anotados entre paréntesis, como el verbo de la traducción de (8).

(8) T: El niño rubio (está) sentado
D: HOMBRE RUBIO SENTAR
I: SENTAR