

## Anotación de corpus para o estudo da expresión gramatical de eventos: notas sobre o deseño do proxecto RADIS

Ania Pérez<sup>1</sup>

José M<sup>a</sup> García-Miguel

Carmen Cabeza

Universidade de Vigo

### RESUMO

O proxecto RADIS pretende basear a descrición gramatical da lingua de sinais española (LSE) en usos contextualizados. En concreto, o obxectivo fundamental é a compilación e análise da estrutura das predicacións en LSE desde unha perspectiva tipolóxico-comparada.

Brevemente móstrase como se fai a anotación en ELAN, en tres fases ou niveis: anotación primaria, de compoñentes non manuais (CNM) e secundaria ou gramatical. Tal anotación, nomeadamente a gramatical, permite analizar os principais recursos para a expresión de argumentos na LSE, clasificados en externos (pronomes ou frases nominais), internos (construcións descritivas con clasificadores e locus) e outros que non se apoian nunha articulación manual determinada, como é o caso da acción construída, baseada en compoñentes non manuais, e outros exemplos de mantemento do referente no discurso sen mención explícita.

A utilización de métodos cuantitativos permite correlacionar as formas de expresión con diferentes xéneros discursivos.

**Palavras-chave:** lingua de sinais española (LSE), lingüística de corpus, anotación, estrutura argumental

### ABSTRACT

The RADIS project aims to approach the grammatical description of Spanish Sign Language (LSE) from contextualized uses. In particular, its main objective is to compile and analyse the LSE predications from a typological-comparative perspective. The process of annotating in ELAN is hereby briefly described. It

---

<sup>1</sup> Email de contacto: [aniaperezperez@uvigo.es](mailto:aniaperezperez@uvigo.es)

consists of three phases or levels: primary annotation, annotation of non-manual components, and secondary or grammatical annotation.

This method—in particular the grammatical annotation—allows us to analyse the main resources for expressing event participants in LSE. Verb arguments are categorised into external (pronouns or nominal phrases), internal (descriptive constructions with classifiers and locus) and those non-based on a specific manual articulation, as it is the case for constructed action, expressed by non-manual components, as well as other examples of referent maintenance throughout the discourse without any explicit mention.

The use of quantitative methods enables the correlation of forms of expression with different types of discourse.

**Keywords:** Spanish Sign Language (LSE), corpus linguistics, annotation, argument structure

## 0 Introducción

A investigación gramatical baseada en corpus ten xa unha certa implantación no que ás linguas de sinais se refire (de Beuzeville, Johnston, & Schembri, 2009; Cormier, Fenlon, & Schembri, 2015; Grabielle Hodge, 2013; Johnston, 2013). O desenvolvemento de programas informáticos que permiten anotar vídeos, nomeadamente ELAN (vid. por exemplo Crasborn & Sloetjes, 2014), favoreceu a posta en marcha de proxectos de elaboración, etiquetado e análise de corpus destas linguas que requiren o manexo da imaxe, desde hai máis dunha década.

O proxecto RADIS enmárcase nesta tradición, malia que os seus obxectivos sexan máis modestos no que atinxe ás dimensións e concepción da mostra<sup>2</sup>, pois non se programou un corpus representativo e estruturado, mais un conxunto de exemplos de discurso en lingua de sinais española (LSE) que permitisen basear a descrición gramatical en usos contextualizados. A compilación e análise da mostra pretende fundamentalmente coñecer a estrutura das predicacións en LSE desde unha perspectiva tipolóxico-comparada.

A utilización de datos reais de discurso permite identificar patróns argumentais preferidos para a expresión gramatical de estados de cousas, e mesmo atopar correlacións entre tipos de expresión segundo o tipo de discurso, como se exemplifica neste traballo. Con este propósito, trabállase sobre unha mostra de gravacións nas que predomina o xénero narrativo, en boa parte porque a pretensión de comparar linguas aconsella utilizar materiais de elicitación de historias (“retelling”) xa usadas

---

<sup>2</sup> O Centro de Normalización Lingüística da LSE (CNLSE) está comprometido na elaboración dun corpus da LSE que reflicta a variación real desta lingua. Ver información en <http://www.cnlse.es/es/corpus-de-la-lengua-de-signos-esp%C3%B1ola> (con acceso 4/9/2018).

para a investigación interlingüística, tales como *Pear Film* (Chafe, 1980) ou *Frog, where are you?* (Mayer, 1969) Por outra banda, a introdución doutros tipos de discurso permite explicar datos que poidan derivarse de aspectos contextuais. A Táboa 1 resume a composición do corpus no momento de escribir este artigo.

Tipo de discurso	Número de gravacións	Duración
Conversa	3	00:11:43
Exemplos elicitados de oracións illadas	3	01:05:47
Entrevistas	9	00:31:15
Narrativa ( <i>Pear Film</i> ; <i>Frog, where are you?</i> ; outros)	21	01:03:11
Webserie ( <i>Mírame cuando te hablo</i> <sup>3</sup> )	5	01:15:14
Outros	4	00:01:28
<b>Total</b>	<b>45</b>	<b>04:08:38</b>

**Táboa 1.** Composición actual do corpus do proxecto RADIS

O corpus non é moi extenso; pero, como queda dito, a finalidade non é contar cun corpus amplo e representativo de toda a lingua, senón dispor de materiais que poidan ser anotados en detalle e analizados en profundidade. Entendemos que a mostra é máis que suficiente para obter algunhas conclusións sobre patróns gramaticais e a súa distribución no discurso, e máis limitada para obter conclusións sobre o léxico da LSE. Neste traballo descríbese brevemente como se desenvolven a segmentación e a anotación no proxecto RADIS, e preséntase a base de datos léxica que complementa o corpus anotado (apartado 1). A seguir explícase, tamén con brevidade, como se resolven algunhas dificultades concretas de identificación e tratamento de unidades léxicas e variantes, así como tamén de recoñecemento de categorías léxicas (sección 2). O apartado 3 focaliza o interese principal do proxecto, que é o estudo da expresión gramatical das relacións eventivas e de como inciden os diferentes xéneros discursivos na selección de determinado tipo de expresión gramatical. O artigo remata cunha breve recapitulación (apartado 4).

## 1 Punto de partida

### 1.1 Anotación

A tarefa de anotación do proxecto RADIS segue as pautas establecidas para o corpus da lingua de sinais australiana (AUSLAN na súa abreviatura sobre o nome en inglés

<sup>3</sup> Cinco capítulos da segunda temporada da serie producida por Idendead co título *Mírame cuando te hablo*.

*Australian Sign Language*) (Johnston, 2010, 2016)<sup>4</sup>. Paralelamente á anotación do corpus desenvólvese unha base de datos léxica que identifica cada un dos ítems anotados (*tokens*, ver infra). Sobre tal modelo, deseñáronse tres fases ou niveis de anotación do corpus RADIS:

- Anotación primaria: segmentación inicial, glosado de unidades léxicais e semi-léxicais, separación de man dereita e man esquerda, tradución ao español, *role shift*. Máis da metade das gravacións que compoñen actualmente o corpus (28/45) teñen completada a anotación primaria.
- CNM (compoñentes non manuais): mirada, movementos na zona dos ollos, xestos da boca, oralizacións, movementos da cabeza, torso e ombros. Na actualidade esta fase está completa para 13 das gravacións do corpus (13/45).
- Anotación secundaria: categorías gramaticais, estrutura argumental, *locus*, animación. Esta fase, a máis complexa, está actualmente realizada para 13 das mostras gravadas do corpus (13/45)

O exemplo (1) mostra un exemplo do primeiro nivel, a anotación primaria.

(1) *Liñas de anotación primaria*

	3.000	00:00:56.500	00:00:57.000	00:00:57.500	00:00:58.500
Ref [58]	VS-CBB#036		VS-CBB#037		
Trad_lit [48]	Me explicaba [algo]		(y) yo lo entendía (muy) claro.		
MD_Glosa [144]	EXPLICAR		INDX. ENTENDER		CLARO
MI_Glosa [144]	EXPLICAR				CLARO
Trad_lib [2]	Entendía con mucha facilidad lo que me explicaba.				

O paso inicial, a segmentación, é crítico, debido a que constitúe unha das primeiras decisións, e está intimamente unido ao problema teórico de que unidades gramaticais é preciso recoñecer. De maneira tentativa, e á espera de desenvolver estudos rigorosos sobre o particular, establécense segmentos de dous tipos:

- *Clause-like units* (CLUs) son grosso modo segmentos que recoñecen un evento cos seus participantes asociados:  
*Clause-like units* are chunks of signing that prompt the conceptualisation of relations between participants and events. They are ‘clause-like’ because we do not yet know if they are ‘clauses’ in the sense of those identified in spoken or written languages, or if they are some other type of construction (Hodge, 2013).

<sup>4</sup> O corpus de AUSLAN serviu tamén de inspiración para a elaboración do corpus da lingua de sinais británica (BSL nas súas siglas en inglés, ver (Cormier et al., 2012).

- *Tokens*: son segmentos equivalentes a palabras. Máis concretamente, representan unha articulación manual (da man dereita, da man esquerda ou de ambas as dúas mans) que idealmente ten unha correspondencia na base de datos léxica, pero pode ser tamén un xesto, un clasificador ou un deíctico.

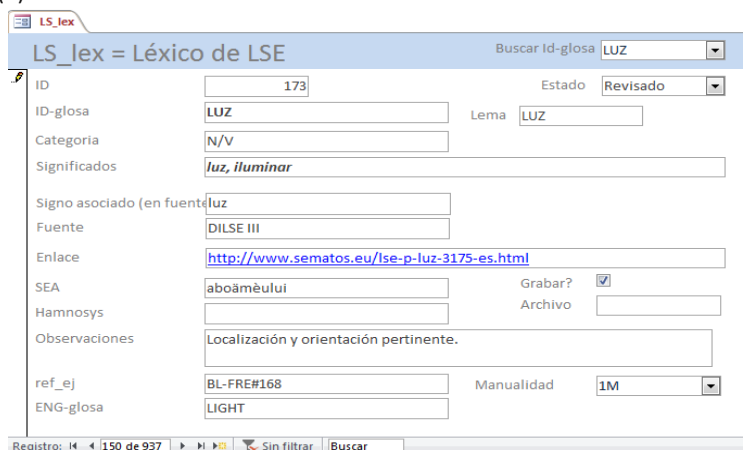
A liña de referencia (Ref) identifica as CLU e as anotacións nas liñas MD\_Glosa e MI\_Glosa recollen os tokens de cada man (exemplo 1).

## 1.2 Base de datos léxica

Como xa se adiantou no apartado anterior, en paralelo á anotación do corpus desenvólvese unha base de datos léxica que contén información de todos os tokens segmentados e identificados. Isto inclúe signos léxicos, índices e algunhas boas. Non están incorporados os xestos que non forman parte do léxico da LSE.

En (2) móstrase como exemplo unha entrada da base de datos léxica. A ID-glosa é a forma que remite a un signo ou a unha variante dun signo. Como se explica máis adiante (apartado 2.1) non coincide exactamente co lema, xa que varias ID-glosas poden remitir a un mesmo lema. O rexistro fornece información sobre o significado e orienta sobre a categoría ou categorías léxicas a que corresponde o signo en cuestión. É importante clarificar a que articulación manual corresponde cada entrada, e por iso se dá información sobre unha fonte externa (se existe) e se vincula o signo cun vídeo creado *ad hoc*. Nalgúns casos, como o do exemplo (2), pode incluírse un enlace cun vídeo. Os campos SEA e HamNoSys teñen a misión de ofrecer unha transcripción fonética en dous sistemas creados especificamente para as linguas de sinais<sup>5</sup>.

### (2) Rexistro da base de datos léxica



The screenshot shows a web interface for the LSE Lexicon. The title is 'LS\_lex = Léxico de LSE'. A search bar at the top right contains 'LUZ'. The main form displays the following fields:

ID	173	Estado	Revisado
ID-glosa	LUZ	Lema	LUZ
Categoría	N/v		
Significados	luz, iluminar		
Signo asociado (en fonte)	luz		
Fuente	DILSE III		
Enlace	<a href="http://www.sematos.eu/lse-p-luz-3175-es.html">http://www.sematos.eu/lse-p-luz-3175-es.html</a>		
SEA	aboãmèului	Grabar?	<input checked="" type="checkbox"/>
Hamnosys		Archivo	
Observaciones	Localización y orientación pertinente.		
ref_ej	BL-FRE#168	Manualidad	1M
ENG-glosa	LIGHT		

At the bottom, there is a navigation bar showing 'Registro: 14', '150 de 937', and a search button.

<sup>5</sup> SEA equivale a *Sistema de Escritura Alfabética* (Herrero Blanco, 2003) e HamNoSys a *Hamburg Notation System for Sign Languages* (Hanke, 2004).

Como se explica máis adiante (apartado 3) outras táboas da base de datos conteñen os verbos do corpus, con cadansúa estrutura argumental. A base de datos de verbos recolle tanto as formas léxicas (DAR, COMER, etc), como os predicados descritivos (os chamados “clasificadores”), que reciben unha anotación específica na que se fai constar o tipo de clasificador de que se trata, a configuración manual utilizada e unha descrición esquemática do significado. Así, no exemplo (3) móstrase como se anota no corpus un predicado descritivo que representa unha persoa que cabalga sobre un cervo. Ambas as mans representan entidades: na dereita os dedos índice e medio estirados e orientados cara abaixo representan a persoa que vai sobre o lombo do animal; na esquerda, a configuración de man estendida en horizontal alude ao cervo.

(3) MD: cl.e(2b):PERSONA-sobre-ciervo  
MI: cl.e(Bh):CIERVO-correr

Cada entrada da base de datos de verbos recolle información sobre a súa estrutura semántica tal e como se deduce das aparicións do corpus. Os argumentos numéranse como A1, A2... e especificase para cada un o significado que aporta ao estado de cousas (micro-rol, ver infra, apartado 3).

## 2 Identificación e categorización de unidades léxicas

### 2.1. Tokenización: ID-glosas e lematización

Os signos léxicos anótanse con glosas identificativas (*ID-glosas*), etiquetas que vinculan as formas do corpus coa base de datos léxica. Unha ID-glosa é unha palabra (ou un conxunto de palabras unidas con guión) dunha lingua vocal coñecida polas persoas xordas usuarias da lingua (no caso da LSE, o español), escrita en letras maiúsculas. A palabra (ou palabras) escollida coincide cun dos significados que ten o sinal en cuestión. Unha ID-glosa remite sistematicamente a un signo aínda que este presente variación (Crasborn & Meijer, 2012; Johnston, 2010).

A tarefa de distinguir ID-glosas complícase cando é preciso facer fronte a decisións sobre que debe recoñecerse como una forma léxica diferente ou, dito doutro modo, que é un lexema e que é unha variante. A solución pasa por establecer diferentes niveis nas posibles variacións dunha forma: 1) variacións previsibles tanto na forma como no significado; 2) variacións atribuíbles a modificacións dun mesmo lexema; 3) variacións que remiten a lexemas diferentes.

As diferencias de forma que levan aparelladas diferencias de significado de xeito regular e previsible (que poderíamos considerar semellantes aos fenómenos de flexión que dan lugar a paradigmas nas linguas vocais) reciben unha mesma ID-glosa. Isto é, considéranse formas do mesmo signo léxico. Así, por exemplo, as articulacións do verbo DAR que inclúen información sobre Axente e Receptor (por exemplo, se o

movemento e orientación van da persoa signante a outra localizada no espazo de signación ou á inversa, segundo quen dea a quen) non teñen ID-glosas diferentes (nin tampouco lemas diferentes).

No que toca ao segundo tipo, hai tres tipos de variación que producen a creación de máis dunha ID-glosa do mesmo signo: variación fonética, incorporación numeral e manualidade.

- A variación fonética consiste na substitución dun único parámetro articulatorio con respecto á forma canónica de cita e non ten consecuencias no significado. Por exemplo, o signo PADRE (“pai”) en LSE pode variar na configuración da man. As ID-glosas correspondentes son PADRE e PADRE(2p) para dar conta da variación.

(4) a. PADRE



b. PADRE(2p)



- Polo que respecta á incorporación numeral, é un proceso morfolóxico que afecta a un conxunto limitado de signos léxicos que cambian a súa configuración ao combinarse con numerais. Por exemplo, para expresar o significado “dentro dun ano” utilízase en LSE o signo glosado como AÑO.FUTURO. Se se quere dicir “dentro de dous anos” este modifica a súa configuración de xeito que o resultado ten a ID-glosa AÑO.FUTURO(DOS).

(5) a. AÑO.FUTURO



b. AÑO.FUTURO(DOS)



- Por manualidade entendemos variación no número de articuladores manuais (articulación monomanual ou bimanual) ou na relación de dominancia entre eles reflectida no movemento. Así, por exemplo, o verbo DAR pode realizarse con dúas mans, sendo o caso que a forma canónica é monomanual. Nese caso a ID-glosa é DAR(2M).

(6) a. DAR



b. DAR(2M)



Nestes tres casos mantense a referencia sistemática ao signo a través do lema, que é común ás ID-glosas que representan as variantes. O lema correspóndese cunha palabra en máiusculas da lingua española (ou máis dunha).

As formas que se diferencian en máis dun parámetro articulatorio, aínda que o significado sexa próximo, serán consideradas lexemas diferentes, nomeadamente cando a diferenza no significado non é regular nin previsible. Por exemplo, na mostra de LSE recóllense os lexemas a que se atribúen as ID-glosas COMER e COMER2. Malia a súa semellanza no significado (“comer” vs. “engulir”) constitúen lemas diferentes porque o que as distingue non é algo regular nin previsible.

## 2.2. Categorias léxicas

Tanto as unidades léxicas como as súas variantes poden ser polifuncionais, isto é, entendemos que un lexema pode ser categorialmente flexible, e expresar significados que noutras linguas categorizamos como nomes, verbos ou adxectivos. Na nosa análise da LSE a categorización morfosintáctica non é por si soa un criterio para separar lemas diferentes. Por exemplo, o signo identificado pola ID-glosa BICICLETA, tipicamente bimanual aínda que cabe a posibilidade de encontrala como monomanual, podemos traducila, dependendo do contexto, como “bicicleta” ou como “andar en bicicleta / pedalear”. No primeiro caso equivale a un nome e no segundo a un verbo, pero na nosa base de datos léxica é un único lema categorizado como “N/V”. Sen embargo, na anotación secundaria do corpus si é necesario indicar cal é o sentido e as relacións sintáctico-semánticas específicas que se establecen nese contexto, e nas liñas de anotación correspondentes indícase a categoría particular apropiada a ese contexto. Outros exemplos similares son os de ROPA (= “roupa,



vestir” ), SIGNAR (“signar, lingua de sinais”), FOTOGRAFÍA (“fotografía, fotografar, fotográfico”)

Lema	Categoría	Significados
BICICLETA	N/V	bicicleta, pedalear, andar en bicicleta
ROPA	N/V	roupa, vestir
SIGNAR	N/V	signar, lingua de sinais, falar
FOTOGRAFÍA	N/V/Adx	fotografía, fotografar, fotográfico

**Táboa 2.** Exemplos de lexemas con categoría flexible no corpus de LSE

O sistema de anotación permite recuperar de maneira automática a información categorial contida na base de datos léxica para cada token anotado en MD\_Glosa e engadila á liña MD\_Cat (ver exemplo 5), e a continuación realízase manualmente a desambiguación categorial escollendo a categoría específica válida para cada contexto particular.

### 3 A estrutura argumental e a expresión gramatical de eventos

#### 3.1 Verbos e argumentos

O principal obxectivo da anotación secundaria (morfosintáctica) no proxecto RADIS é a recollida de datos contextualizados sobre a estrutura argumental, a relación entre predicado e argumentos na LSE que poidan ser comparables con datos doutras linguas. A pregunta básica para o deseño do sistema de anotación de relacións sintáctico-semánticas é a de como comparar linguas estruturalmente diferentes dadas as diferenzas estruturais entre unhas e outras, diferenzas que son aínda maiores se comparamos linguas de modalidades diferentes. Na nosa opinión a base de comparación debe situarse na semántica e no discurso, na capacidade de calquera lingua de expresar calquera contido. Precisamente, a construción do noso corpus a partir de estímulos visuais permítenos observar os recursos expresivos dispoñibles na LSE para narrar unha breve historia e para describir eventos e as entidades que participan neles.

Ao segmentar o texto signado en unidades predicativas (CLUs) buscamos identificar dentro delas os tokens que expresan predicados e argumentos, e a continuación etiquetalos respectivamente como V e como A. Como o número de argumentos de cada predicado é variable, numeramos os argumentos tomando como criterio o seu valor semántico e non a orde de aparición (a diferenza do que se fai no corpus Auslan). Por exemplo, nas cláusulas cun predicado como “regalar”, sistematicamente A1 é a persoa que regala, A2 é a cousa regalada e A3 é o receptor do regalo, sexan cales sexan os argumentos expresados explicitamente e a súa orde no texto. A anotación gramatical do corpus, por medio de ELAN, complementase cunha base de datos de predicados e argumentos, onde se clasifican semanticamente os predicados e

defínese o valor semántico de cada argumento (ver supra, § 1.2). Para poder establecer comparacións con outras linguas, esta base de datos está vinculada aos datos da base de datos tipolóxica ValPaL (Hartmann, Haspelmath, & Taylor, 2013) a coa base de datos ADESSE de verbos e construcións verbais do español (García-Miguel, Vaamonde, & González Domínguez, 2010).

En certo sentido, A1, A2, etc. representan roles específicos de cada verbo ou predicado, isto é, ‘micro-roles’ (Hartmann, Haspelmath, & Cysouw, 2014), pero tamén poden entenderse nun senso máis xeral. O A1 como o primeiro argumento entendido como actor protagonista, tipicamente axente, máis ou menos equivale ao que nalgúns linguas poden identificarse como suxeito gramatical. O A2 pode entenderse como obxecto, tipicamente paciente, en predicados biargumentais transitivos, e como localización de referencia con predicados locativos e direccionais.

### 3.2 Maneiras de expresar os argumentos

Unha anotación sintáctico-semántica que queira dar conta do significado expresado por unha expresión non pode limitarse a etiquetar as relacións que manteñen os nomes cos verbos, debe atender tamén aos recursos formais utilizados. Os argumentos semánticos dun predicado (os participantes nos eventos evocados polo predicado) non sempre son expresados mediante nomes ou pronomes que acompañan secuencialmente a ese predicado. Con carácter xeral, nas linguas do mundo é posible expresar os participantes dos eventos tanto de maneira externa ao predicado, como de maneira interna mediante a morfoloxía verbal. Por exemplo, o(s) participante(s) no evento de xogar é expresado mediante un nome en (7a), un pronome en (7b), pero só mediante a flexión verbal en (7c)

- (7) a. O neno xoga  
b. El xoga  
c. Xogamos

Por outro lado, tamén é posible aludir a participantes nos eventos mediante a comunicación non verbal, e no discurso continuo moitas veces un participante nun evento non é expresado de ningún xeito porque é perfectamente identificable no contexto.

Nas linguas xestuais como a LSE, xunto á concatenación secuencial de signos lingüísticos, hai que engadir outras posibilidades expresivas: a utilización do espazo en predicados indicadores (tamén chamados “direccionais”), a utilización de predicados descritivos con configuracións manuais clasificadoras, que poden aludir a entidades diferentes en cada man, a expresión non manual de accións dos participantes no evento designado (no que se coñece como acción construída e cambio de rol), etc.

### 3.2 A anotación secundaria no corpus RADIS

O noso sistema de anotación en ELAN prevé liñas específicas para cada unha destas posibilidades especificadas ao final do apartado anterior. As liñas que aparecen en cor amarela no exemplo (8) son as que corresponden á anotación secundaria ou anotación gramatical.

(8) *Anotación secundaria no corpus RADIS*

	000	00:02:02.500	00:02:03.000	00:02:03.500	00:02:04.000
Trad_lib	El niño lo miró molesto, era un búho.				
Ref	HR-ARS#041				HR-ARS#042
Trad_lit	(El niño) miró hacia arriba,				era un búho.
MD_Glosa	MIRAR				RAPAZ^VOLAR
MD_Cat	V.Dir				V
MD_Arg	V				V_A1
MD_Locus	>ia(>A2)				
MD_Anim					An
MI_Glosa					RAPAZ^VOLAR
MI_Cat					
MI_Arg					
MI_Anim					
Rol	AC:niño				
Rol_Arg	A1				



[Vídeo 1](#)

Entre elas, as liñas que sinalan directamente ás formas de expresar a relación entre verbo e argumentos son as recollidas na Táboa 3

MD_Arg	V, A1, A2, ... verbo e argumentos expresados mediante signos na man dereita (dominante)
MI_Arg	V, A1, A2, ... id. expresados mediante signos na man esquerda (se son distintos do expresado simultaneamente na dereita)
Rol_Arg	V, A1, A2, ... expresados non manualmente mediante cambio de rol do narrador
Locus	Localización de argumentos sinalados mediante predicados indicadores / direccionais

**Táboa 3.** Liñas de anotación secundaria que sinalan a estrutura argumental

Complétanse con MD\_Cat e MI\_Cat, que especifican a categoría léxica atribuída a cada token anotado na correspondente liña nai MD\_Glosa ou MI\_Glosa, e con outras dúas (MD\_Anim e MI\_Anim) que especifican a animación dos argumentos expresados nunha e outra man. No exemplo (8) non aparece nada nas liñas correspondentes á categoría, argumentos e animación da man esquerda debido a que a articulación desta man non constitúe un argumento diferente do expresado na man dereita. Si habería que especificar esta información nun caso como o do exemplo (9), que reproduce con máis detalle o anterior exemplo (3). Nel cada unha das mans representa un evento no que intervéñ unha entidade: a man dereita especifica unha localización da persoa sobre o cervo e a man esquerda mostra a carreira do cervo que leva no lombo á persoa.

- (9) MD\_Glosa: cl.e(2b):PERSONA-sobre-cervo  
 MD\_Cat: V.D  
 MD\_Arg: V\_A1  
 MD\_Anim: Hu  
 MI\_Glosa: cl.e(Bh):CIERVO-correr  
 MI\_Cat: V.D  
 MI\_Arg: V\_A1  
 MI\_Anim: An

Trad: *e o cervo comezou a cabalgar co neno no lombo*

Ref: HR-ARS#050



[Video 2](#)

Vemos, pois, que o sistema de anotación prevé que un mesmo token exprese simultaneamente predicado e argumento (no exemplo, V\_A1). Tamén se dá o caso de que un predicado implique un argumento recuperable no contexto pero que carece de forma de expresión explícita (iría anotado como V\_ºA1).

**3.3 A expresión dos argumentos en LSE**

En LSE caben os seguintes tipos de recursos expresivos para participantes en eventos:

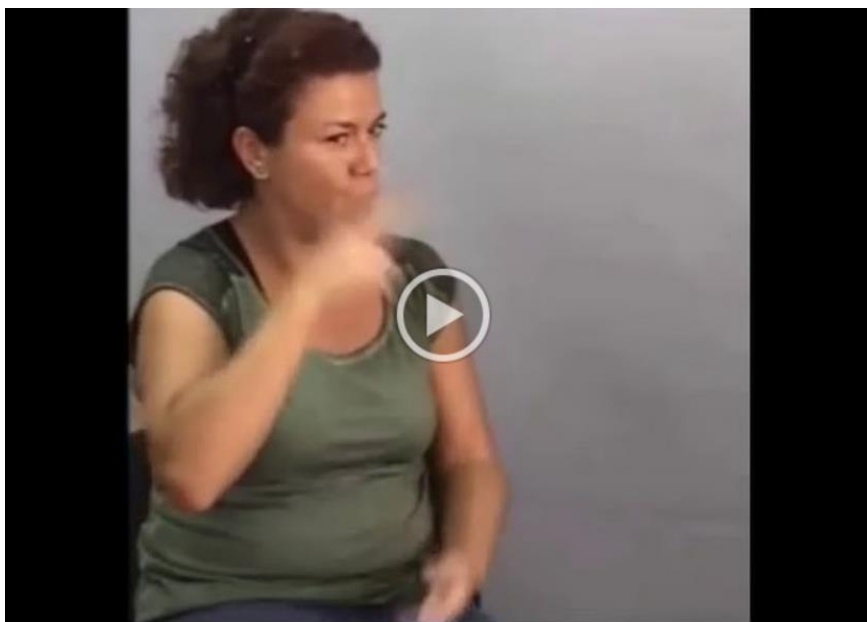
- A. Expresión **externa** ao predicado, mediante léxico nominal [LEX] ou mediante índices deícticos (pronomes) [PRO]

LEX: Léxico

PRO: Índices (~pronomes / deícticos)

(10) Expresión externa

MD_Glosa:	INDX.PRO	VIVIR	DÓNDE VIGO
MD_Cat:	Pro	V.P	Interr N
MD_Arg:	A1	V	A2
MI_Glosa:		VIVIR	
Trad_lit:	<i>Eu vivo en Vigo</i>		
Ref:	ET-MJC#059		



[Video 3](#)

Os dous argumentos do verbo VIVIR (bimanual) están expresados, respectivamente, por un deíctico pronominal que alude á propia signante (A1 na liña MD\_Arg), e por unha frase con valor locativo introducida por un interrogativo (A2). O uso do interrogativo constitúe un recurso de focalización habitual na LSE.

- B. Expresión **interna** no predicado, ben mediante configuracións manuais clasificadoras en predicacións descritivas [CL], ben mediante predicados indicadores/direccionais que sinalan *loci* asociados con participantes [LOC]  
CL: “Clasificadores” (en predicacións descritivas), como no exemplo (9)  
Loc: “Locus” (cara ao que se orientan “verbos direccionais”), como no exemplo (8)

Como se ve no exemplo (9), na carreira do cervo co neno cabalgando sobre o seu lombo, cada unha das mans toma unha forma específica que representa a posición ou o movemento dunha entidade. Trátase, por tanto, de construcións predicativas de carácter descritivo nas que o argumento do que se describe a posición ou o movemento vai inserido no mesmo verbo a través da configuración manual. Ademais, neste caso cada man describe unha acción, aínda que están relacionadas e son simultáneas, unha a carreira do cervo levando o neno no lombo, e outra a posición do neno sobre o lombo do cervo. Este recurso de mantemento da referencia no

discurso está amplamente estudado para as linguas de sinais, desde diferentes perspectivas teóricas e con distintas denominacións: clasificador (*classifier*) ou predicado clasificador (*classifier predicate*) (Aarons & Morgan, 2003; Barberà & Quer, 2018; Benedicto, Cvejanov, & Quer, 2007; Schick, 1990), verbos de movemento e localización (*verbs of motion and location*) (Supalla, 1986), sinais policomponenciais (*polycomponential signs*) (Morgan & Woll, 2007; Slobin et al., 2003) ou predicados / construcións descritivos ou simplemente descrições (*depicting predicates / constructions / depictions*) (Beal-Alvarez & Trussell, 2015; Ferrara & Halvorsen, 2018; Liddell, 2000). É característico destas construcións o feito de estaren só parcialmente sedimentadas ou lexicalizadas, é dicir, combinan elementos que son regulares e esquemáticos (no caso do exemplo, a configuración das mans, que forma parte dun repertorio limitado de unidades) con outros máis idiosincráticos, e ademais son unha fonte de creación de léxico novo. Para o caso que nos ocupa, é relevante a función de mantemento da referencia que desempeña a configuración manual. Apunta a un argumento semántico, que xa foi mencionado con anterioridade no discurso, a través dun recurso interno ao predicado.

No caso do verbo MIRAR, no exemplo (8), o movemento oríentase cara á posición do argumento que representa a cousa mirada. No caso do exemplo, o argumento A2 está localizado á esquerda e arriba da signante, e esa é a orientación que toma a man dominante. MIRAR forma parte dun conxunto de verbos que ven modificada a súa articulación en función do lugar (*locus*) onde se localizan, dentro do espazo de signación, os micro-roles de “mirador” (xenericamente o Axente) e “mirado” (ou Obxecto) (Padden, 1988). É frecuente na LSE (e noutras linguas de sinais) que o locus do Axente se faga coincidir co corpo da persoa signante, pois esta representa o rol do Axente (de xeito que se produce un desprazamento das referencias de persoa, xa que o “el” do neno do exemplo exprésase como un “eu”). Tamén é habitual que a dirección do ollar acompañe esa orientación manual (Hosemann, 2011; Thompson, Emmorey, & Kluender, 2006). Unha parte importante da investigación sobre este tipo de verbos incide na función de concordancia que exercerían estes índices incorporados na articulación manual (e da mirada) (ver, entre outros, (Cormier, Fenlon, et al., 2015; Costello, 2015; Liddell, 2011; Lillo-Martin & Meier, 2011; Meir, 1999; Schembri, 2018).

- C. **Outros:** Expresión non manual de acción construída e cambio de rol de xeito que os xestos do signante evocan un participante no proceso [AC], e alusión a un participante recuperable no contexto pero non expresado formalmente [Ø]

AC: Acción construída-Cambio de rol (signante=personaxe)

Ø: Anáfora cero (de elementos recuperábeis en contexto)

Outro fenómeno amplamente descrito para as linguas de sinais é o que se coñece como acción construída (*constructed action*), cambio de perspectiva ou cambio de rol (*role shift*) (Cormier, Smith, & Sevcikova-Sehyr, 2015; Janzen, 2004; Perniss, 2007). A

persoa signante pode adoptar unha perspectiva obxectiva (*observer perspective*) ou ben tomar o rol da persoa de quen está a falar (*character viewpoint*). No segundo caso, un dos efectos no discurso é que ese uso da perspectiva da personaxe, que se expresa fundamentalmente a través dos compoñentes non manuais (Cormier, Smith, et al., 2015), é a persistencia dun dos argumentos do verbo (Cabeza & García-Miguel, en prensa). É por tanto un recurso que explota a corporeización da referencia a un argumento e enténdese que facilite a non aparición da mención expresa a tal personaxe. No exemplo (11) o signante está asumindo o rol do lobo que está a soprar sobre a casa dos tres porquiños. Vese que a súa expresión facial reflicte o esforzo que o lobo está a facer:

(11) Acción construída

MD_Glosa:	SOPLAR(2M)
MD_Cat:	V
MD_Arg:	V_ºa1
Rol:	AC: lobo
Trad_lit::	Sopla (intensamente)
Ref:	HC-JRV#112



[Video 4](#)

Por último, cómpre aludir a outras maneiras máis sutís de facer referencia aos obxectos do discurso. Hodge, Anible e Ferrara (2018) chaman a atención sobre certas estratexias semióticas que desenvolven as persoas signantes para facer inferir elementos no discurso e manter a súa referencia a través de procedementos que chaman “visible/invisible surrogates”: os autores advirten da conveniencia de



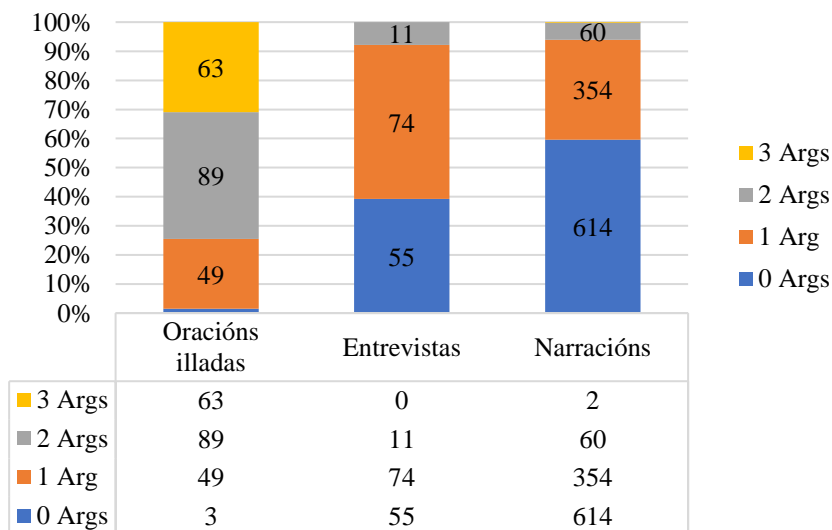
diferenciar estes recursos da mera anáfora cero, isto é a non mención explícita dun argumento por ser este recuperable contextualmente.

### 3.4 Recursos expresivos e xénero discursivo

Como era de esperar, os diferentes recursos expresivos non se distribúen uniformemente no discurso, e a distribución depende de diversos factores sintácticos, semánticos e pragmáticos.

A Figura 1 mostra o número de argumentos “externos” dependendo do xénero discursivo. O número de argumentos (a valencia) depende do significado do predicado; pero só en oracións illadas atopamos os argumentos previstos pola valencia verbal expresados mediante elementos léxicos nominais. No discurso continuo, e máis especificamente no discurso narrativo, a maioría dos verbos non van acompañados de nomes na mesma cláusula. En discurso narrativo priorízase pois a expresión interna (clasificadores, indexación), o cambio de rol, ou a non expresión (expresión cero) porque o referente é accesible no contexto.

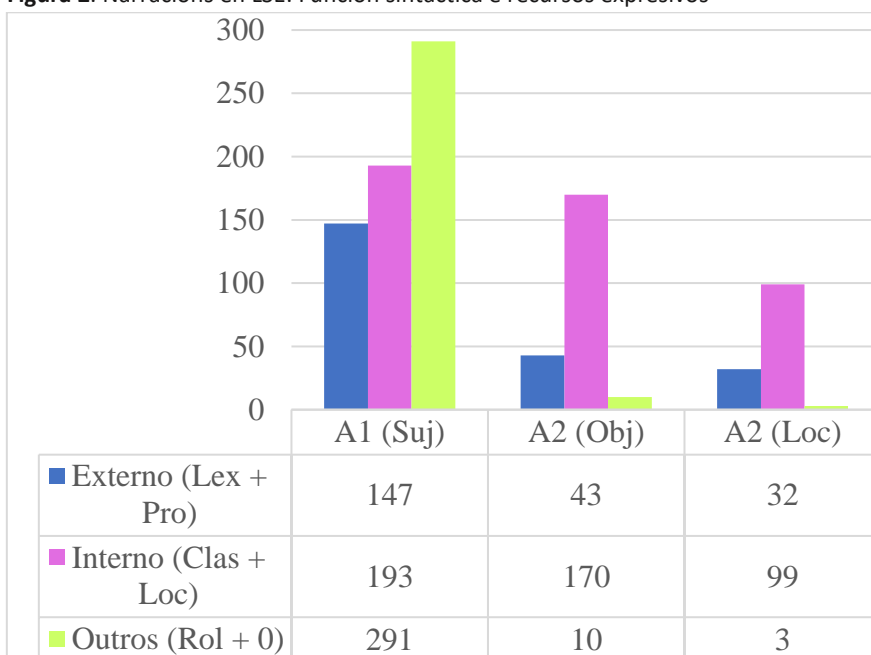
**Figura1.** Número de argumentos “externos” e xénero discursivo



Entre os factores sintácticos que poden determinar a forma de expresión dun argumento, o máis relevante é a función sintáctica. Grosso modo, podemos considerar o primeiro argumento (A1) de cada predicado como equivalente do suxeito, e o segundo argumento (A2) como equivalente de Localización (Loc) con predicados de movemento e localización, e como equivalente de Obxecto (Obx) con outros predicados biargumentais. Teoricamente, esperaríamos menor frecuencia na expresión léxica do axente de predicados biargumentais (Du Bois, 1987; Du Bois,

Kumpf, & Ashby, 2003) Acabamos de ver na Táboa 4 que nas narracións é rara a expresión dos argumentos mediante elementos léxicos ou pronominais externos ao predicado; pero, como se pode observar na Figura 2, na LSE tanto para A1 como para A2 recórrase máis frecuentemente á expresión interna mediante configuracións clasificatorias ou á indexación. No que si é claramente diferente o comportamento do suxeito (A1) é na asociación co cambio de rol ou na omisión de referentes accesibles, o que mostra a súa asociación non só coa información dada senón tamén co punto de vista ou empatía (Kuno & Kaburaki, 1977)

**Figura 2.** Narracións en LSE: Función sintáctica e recursos expresivos



#### 4 Recapitulación

Preséntase o proxecto RADIS, cuxo obxectivo é basear a descrición gramatical da LSE en usos contextualizados. En concreto, a compilación e análise da mostra pretende fundamentalmente coñecer a estrutura das predicacións en lingua de sinais española (LSE) desde unha perspectiva tipolóxico-comparada.

Brevemente móstrase como se fai a anotación en ELAN, en tres fases ou niveis: anotación primaria, de compoñentes non manuais (CNM) e secundaria ou gramatical. Tal anotación, nomeadamente a gramatical, permite analizar os principais recursos para a expresión de argumentos na LSE, clasificados en externos (pronomes ou frases nominais), internos (construcións descritivas con clasificadores e locus) e outros que

non se apoian nunha articulación manual determinada, como é o caso da acción construída, baseada en compoñentes non manuais, e outros exemplos de mantemento do referente no discurso sen mención explícita.

A utilización de métodos cuantitativos permite correlacionar as formas de expresión con diferentes xéneros discursivos.

Pero o uso de métodos propios da lingüística de corpus pode estenderse ao estudo de moitos outros fenómenos gramaticais e discursivos, e facilita a investigación interlingüística, nomeadamente se se gravan narracións comparables sobre estímulos comúns como *Pear Film* ou *Frog, where are you?*.

## 5 Agradecementos

Gobierno de España, *Ministerio de Ciencia, Investigación y Universidades* (Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia: FFI2013-41929-P & FFI2016-78194-P)

Xunta de Galicia, *Consellería de Cultura, Educación e Ordenación Universitaria* (R2014/007 & ED431D 2016/011)

Colaboraron: Juan Ramón Valiño, Francisco Eijo, Cristina Freire, Fernando J. García

## Referencias

- Aarons, D., & Morgan, R. (2003). Classifier predicates and the creation of multiple perspectives in South African sign language. *Sign Language Studies*, 3(2), 125-156.
- Barberà, G., & Quer, J. (2018). Nominal referential values of semantic classifiers and role shift in signed narratives. En A. Hübl & M. Steinbach (Eds.), *Linguistic Foundations of Narration in Spoken and Sign Languages* (pp. 251-274). Amsterdam: John Benjamins.
- Beal-Alvarez, J. S., & Trussell, J. W. (2015). Depicting Verbs and Constructed Action: Necessary Narrative Components in Deaf Adults' Storybook Renditions. *Sign Language Studies*, 16(1), 5-29.
- Benedicto, E., Cvejanov, S., & Quer, J. (2007). Valency in classifier predicates: A syntactic analysis. *Lingua*, 117(7), 1202-1215.
- de Beuzeville, L., Johnston, T., & Schembri, A. (2009). The use of space with indicating verbs in Auslan: A corpus-based investigation. *Sign Language & Linguistics*, 12(1), 53-82.
- Cabeza, C., & García-Miguel, J. M. (2018). Verbalización de eventos de movimiento y localización en lengua de señas española (LSE): un análisis de las construcciones descriptivas con clasificador de entidad. *Onomázein*, 41, 227-263.
- Chafe, W. L. (Ed.). (1980). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, N.J.: Ablex.

- Cormier, K., Fenlon, J., Johnston, T., Rentelis, R., Schembri, A., Rowley, K., Adam, R., et al. (2012). From Corpus to Lexical Database to Online Dictionary: Issues in Annotation of the BSL Corpus and the Development of BSL SignBank. *Proceedings of LREC 2012* (pp. 7-12). Apresentado en 5th Workshop on the Representation of Sign Languages: Interactions between Corpus and Lexicon-LREC, Estambul.
- Cormier, K., Fenlon, J., & Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics*, *1*(1). Recuperado febrero 8, 2016, a partir de <http://www.degruyter.com/view/j/opli.2014.1.issue-1/opli-2015-0025/opli-2015-0025.xml>
- Cormier, K., Smith, S., & Sevcikova-Sehyr, Z. (2015). Rethinking constructed action. *Sign Language & Linguistics*, *18*(2), 167-204.
- Costello, B. (2015). *Language and modality: Effects of the use of space in the agreement system of lengua de signos española (Spanish Sign Language)* (Tesis doctoral). Universiteit van Amsterdam & Euskal Herriko Unibertsitatea. Recuperado a partir de [http://www.lotpublications.nl/Documents/415\\_fulltext.pdf](http://www.lotpublications.nl/Documents/415_fulltext.pdf)
- Crasborn, O., & de Meijer, A. (2012). From corpus to lexicon: the creation of ID-glosses for the Corpus NGT. *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon* (pp. 13-17). Apresentado en Language Resources and Evaluation Conference.
- Crasborn, O., & Sloetjes, H. (2014). Improving the exploitation of linguistic annotations in ELAN. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (pp. 3604-3608). Apresentado en 9th International Conference on Language Resources and Evaluation, Reykjavik.
- Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, *63*(4), 805-855.
- Du Bois, J. W., Kumpf, L. E., & Ashby, W. J. (Eds.). (2003). *Preferred argument structure: Grammar as architecture for function*. Amsterdam: John Benjamins.
- Ferrara, L., & Halvorsen, R. P. (2018). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture*, *16*(3), 371-395.
- García-Miguel, J. M., Vaamonde, G., & González Domínguez, F. (2010). ADESSE, a database with syntactic and semantic annotation of a corpus of Spanish. *LREC 2010 - Proceedings of the Seventh International Conference on Language Resources and Evaluation* (pp. 1903-1910). Valetta (Malta): ELRA.
- Hanke, T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. *LREC 2004, Workshop proceedings: Representation and processing of sign languages* (pp. 1-6). Apresentado en LREC.
- Hartmann, I., Haspelmath, M., & Cysouw, M. (2014). Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language*, *38*(3), 463-484.

- Hartmann, I., Haspelmath, M., & Taylor, B. (Eds.). (2013). *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Recuperado a partir de <http://valpal.info/>
- Herrero Blanco, Á. L. (2003). *Escritura alfabética de la lengua de signos española: once lecciones*. Textos docentes. Alicante: Publ. de la Univ. de Alicante.
- Hodge, Gabrielle, Anible, B., & Ferrara, L. (2018, julio 30). *The in/visiblence of doing reference in a Deaf signed language*. Presentado en Sign CAFE 1, Birmingham.
- Hodge, Grabielle. (2013). *Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language)* (Ph.D. thesis). Macquarie University, Sydney.
- Hosemann, J. (2011). Eye gaze and verb agreement in German Sign Language: A first glance. *Sign Language & Linguistics*, 14(1), 76-93.
- Janzen, T. (2004). Space rotation, perspective shift, and verb morphology in ASL. *Cognitive Linguistics*, 15(2), 149-174.
- Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1), 106-131.
- Johnston, T. (2013). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of «pronouns» in Auslan? *Corpus Linguistics & Linguistic Theory*, 9(1), 109-159.
- Johnston, T. (2016). Auslan Corpus Annotation Guidelines. Recuperado a partir de [http://media.auslan.org.au/attachments/Auslan\\_Corpus\\_Annotation\\_Guidelines\\_November2016.pdf](http://media.auslan.org.au/attachments/Auslan_Corpus_Annotation_Guidelines_November2016.pdf)
- Kuno, S., & Kaburaki, E. (1977). Empathy and Syntax. *Linguistic Inquiry*, 8(4), 627-672.
- Liddell, S. K. (2000). Indicating verbs and pronouns: Pointing away from agreement. En K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited* (pp. 303-320). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Liddell, S. K. (2011). Agreement disagreements. *Theoretical Linguistics*, 37(3/4), 161-172.
- Lillo-Martin, D., & Meier, R. P. (2011). On the linguistic status of 'agreement' in sign languages. *Theoretical Linguistics*, 37(3-4), 95-141.
- Mayer, M. (1969). *Frog, Where Are You?* New York: Dial Press.
- Meir, I. (1999). Thematic structure and verb agreement in Israeli Sign Language. *Sign Language & Linguistics*, 2(2), 263-270.
- Morgan, G., & Woll, B. (2007). Understanding sign language classifiers through a polycomponential approach. *Lingua*, 117(7), 1159-1168.
- Padden, C. (1988). *Interaction of morphology and syntax in American sign language*. Outstanding dissertations in linguistics. New York: Garland.
- Perniss, P. (2007). Locative functions of simultaneous perspective constructions in German Sign Language narratives. En M. Vermeerbergen, L. Leeson, & O.

- Crasborn (Eds.), *Simultaneity in signed languages: Form and function* (pp. 27-54). Amsterdam: John Benjamins.
- Schembri, A. (2018). *Indicating verbs and cognition-to-form mapping: pointing away from «agreement» and back again?* Apresentado em Sign CAFE 1, Birmingham.
- Schick, B. S. (1990). Classifier predicates in American Sign Language. *International Journal of Sign Linguistics*, 1(1), 15-40.
- Slobin, D. I., Hoiting, N., Kuntze, M., Lindert, R., Weinberg, A., Pyers, J., Anthony, M., et al. (2003). A Cognitive/Functional Perspective on the Acquisition of “Classifiers”. En K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 271-296). Mahwah, N.J: Lawrence Erlbaum Associates.
- Supalla, T. R. (1986). The classifier system in American Sign Language. En C. G. Craig (Ed.), *Noun classes and categorization: proceedings of a Symposium on Categorization and Noun Classification, Eugene, Or., Oct. 1983* (pp. 181-214). Amsterdam: John Benjamins.
- Thompson, R. L., Emmorey, K., & Kluender, R. (2006). The relationship between eye gaze and verb agreement in American Sign Language: An eye-tracking study. *Natural Language & Linguistic Theory*, 24(2), 571-604.