

HACIA UN TREEBANK DE DEPENDENCIAS PARA LA LSE

JOSÉ M. GARCÍA-MIGUEL Y CARMEN CABEZA

Universidade de Vigo

Title: Towards a dependency treebank for Spanish Sign Language (LSE)

Abstract: This paper presents the foundations, procedures, tests and first results of a dependency treebank of the Spanish Sign Language (LSE). Dependency syntax offers many advantages over other alternatives for the systematic and exhaustive syntactic analysis of a corpus. Nevertheless, the visual modality that is characteristic of sign languages poses unique challenges for their syntactic analysis, among which the most prominent is the simultaneity of expression: both hands, face and other non-manual components. Taking into account these and other particularities of sign languages, the paper explores the main difficulties faced when one tries to apply some usual categories and relations from the syntactic analysis of spoken and written languages to LSE.

Key words: Corpus. Dependency Grammar. Spanish Sign Language. Syntax.

1. INTRODUCCIÓN

Todo estudio científico necesita datos objetivos en los que basarse y en el estudio del lenguaje y las lenguas los datos están constituidos principalmente por los usos verbales de los hablantes. En las últimas décadas, la Lingüística de Corpus ha desarrollado procedimientos para compilar, catalogar, anotar y explorar textos orales y escritos que puedan constituir muestras representativas de una lengua o variedad de lengua. Al mismo tiempo, se han desarrollado procedimientos y estándares de anotación de las propiedades lingüísticas de los textos en los diferentes niveles de análisis. Para las lenguas más extendidas, como el inglés o el español, existen corpus que cuentan con miles de millones de palabras, y también muchos corpus y otros recursos con anotación morfosintáctica, sintáctica, semántica, etc. de los textos que los constituyen. Sin embargo, de muchas otras lenguas del mundo en comparación apenas disponemos de datos en los que basar las descripciones. En particular, son muy pocas las lenguas de signos para las que existe algún tipo de corpus y menos aún aquellas para las que los corpus existentes cuentan con algún tipo de anotación gramatical. Y es muy importante que existan recursos de este tipo, no solo porque cada sistema lingüístico es único sino

también porque las lenguas de signos, al utilizar una modalidad diferente (el canal visuogestual) presentan características estructurales únicas.

El objetivo de este trabajo es contribuir al desarrollo de materiales primarios para el estudio de las lenguas de signos. Lo que presentamos en este artículo es una prueba piloto de anotación sintáctica de un pequeño corpus de lengua de signos española (LSE), con el objetivo de evaluar los principales problemas de anotación con vistas a la constitución de un corpus anotado sintácticamente (un *treebank*) de LSE. Las relaciones sintácticas entre *tokens* del corpus se anotan como relaciones de dependencia, es decir, relaciones asimétricas entre *tokens* individuales, por lo que son más fáciles de anotar y de computar que las estructuras sintácticas basadas en constituyentes. El artículo se organiza de la siguiente manera: en el apartado 2 presentamos los principales antecedentes en la sintaxis de dependencias y en la anotación sintáctica de corpus de alguna lengua de signos; en el apartado 3 describimos y ejemplificamos los aspectos más relevantes del procedimiento formal de anotación; en el apartado 4 se examinan algunas características estructurales de las lenguas de signos que dificultan la aplicación de procedimientos de análisis sintáctico inicialmente diseñados para lenguaje oral o escrito; en el apartado 5 se describen y ejemplifican las principales relaciones de dependencia registradas en el corpus en la prueba piloto.

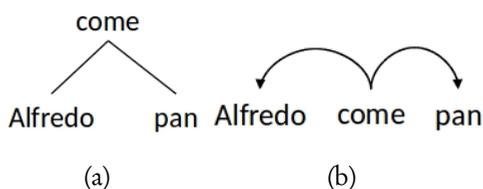
2. ANTECEDENTES

2.1. *Sintaxis de dependencias*

La sintaxis de dependencias es una perspectiva de análisis lingüístico que parte del supuesto de que la estructura sintáctica consiste primariamente en relaciones binarias asimétricas, llamadas dependencias, entre palabras de modo que una de ellas es la principal o regente y la otra es la dependiente o subordinada. Estas estructuras pueden visualizarse en forma de árbol cuyos nodos son palabras y las ramas representan relaciones de dependencia (de Marneffe y Nivre, 2019: 198). Por ejemplo, la estructura sintáctica de *Alfredo*

come pan puede representarse como en el árbol de (a) o como en el diagrama equivalente de (b), indicando así que los nombres *Alfredo* y *pan* dependen del verbo *come*.

(1)



Los antecedentes de la gramática de dependencias pueden buscarse bastante atrás en la tradición gramatical; pero para la lingüística contemporánea la sintaxis de dependencias está vinculada sobre todo a la figura de Lucien Tesnière y su propuesta de sintaxis estructural (Tesnière, 1959). Otros modelos relevantes de sintaxis de dependencias son la teoría Sentido-Texto [*Meaning-Text Theory*, MTT] de Mel'čuk (Mel'čuk, 1988) o el modelo Word Grammar de Hudson (Hudson, 1984).

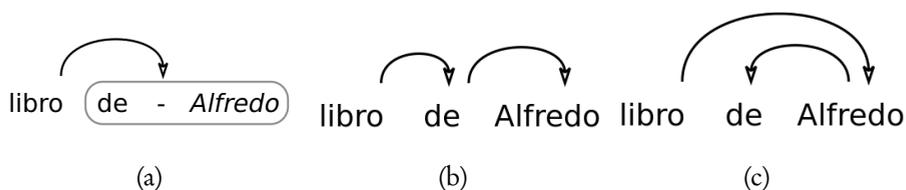
La sintaxis de dependencias no solo se ha asentado en la lingüística teórica y descriptiva contemporánea, sino también en la lingüística computacional (Nivre, 2006). También ha servido de base para la compilación de corpus analizados sintácticamente, o *treebanks*, de los cuales el más conocido es probablemente el *Prague Dependency TreeBank* [PDT] (Böhmová et al., 2003). El proyecto *Universal Dependencies* (<http://universaldependencies.org/>), de aquí en adelante *UD*, proporciona un marco de referencia para la anotación sintáctica de cualquier lengua; que en su versión actual (versión 2.5, de noviembre de 2019) se ha aplicado a 157 *treebanks* de 90 lenguas, entre ellas la lengua de signos sueca, que hasta el momento es la única lengua de signos presente en ese repositorio.

La principal alternativa a la sintaxis de dependencias es el análisis en constituyentes inmediatos, quizás más difundido gracias al distribucionalismo norteamericano (Bloomfield, 1933; Wells, 1947) y sus descendientes generativistas (Chomsky, 1957, 1965, etc.) y también de uso común en algunas propuestas de sintaxis funcional (Longacre, 1970; Rojo y Jiménez Juliá, 1989; Halliday, 2004). Los dos modelos son equivalentes en lo esencial, en el sentido de que casi siempre es posible convertir las representaciones de un modelo al otro sin pérdida de información, pero el modelo dependencial tiene la ventaja de ser más simple al no utilizar frases sino solamente las palabras como unidades básicas entre las que se establecen relaciones. Esta relativa simplicidad también facilita la anotación sintáctica de corpus.

Aun compartiendo las ideas básicas de la sintaxis de dependencia, existen algunas cuestiones controvertidas que han recibido diferentes respuestas en diferentes modelos dependenciales. La primera de ellas es si es posible reducir todas las estructuras sintácticas a relaciones de dependencia. De hecho, Tesnière considera la conexión estructural entre regente y regido como fundamental en sintaxis; pero trata de manera diferente la coordinación, como junción entre elementos del mismo nivel estructural, y los procesos de ‘traslación’ categorial, en los que da un estatus especial a adposiciones y conjunciones subordinantes en tanto que ‘traslativos’ (2)(a). En cambio, la mayoría de los modelos dependenciales posteriores a Tesnière tratan todas las relaciones sintácticas como relaciones de dependencia, esto es, como si todas las construcciones fueran endocéntricas, pero aún difieren en qué elemento debe considerarse núcleo y cuál dependiente especialmente cuando en las conexiones intervienen elementos gramaticales tales como adposiciones, conjunciones o auxiliares, como podemos ver comparando (2)(b) y (2)(c). Algunos de estos problemas pueden resolverse apelando a diferentes niveles de dependencia semántica, sintáctica y morfológica (Mel’čuk, 2003). Al centrarse en un único nivel de análisis, en caso de criterios en conflicto

sobre el núcleo de una conexión, el otorgar prioridad a la conexión directa entre unidades léxicas plenas, como en (2)(c), como se hace en UD, facilita la comparación interlingüística y el paralelismo con la estructura semántica; por lo que en general seguiremos más adelante ese mismo criterio. Para un análisis alternativo, similar al de (2)(b), puede verse la propuesta de Gerdes et al. (2018).

(2) Alternativas en el análisis de relaciones sintácticas en las que interviene una preposición



Las ventajas formales y sustantivas de la sintaxis de dependencias (Mel'čuk, 2003: 219-222), en particular su relativa simplicidad al no requerir categorías frasales, su vinculación con la estructura semántica, y su adecuación para la anotación de corpus y para la comparación interlingüística (Croft et al., 2017), hacen del modelo dependencial el más apropiado también para la anotación sintáctica de muestras de lenguas de signos.

2.2. *Corpus de lenguas de signos con anotación sintáctica*

Empieza a existir cierta tradición en la compilación de corpus de lenguas de signos, y existen corpus relativamente amplios de lenguas como Auslan (Johnston y Schembri, 2006), DGS (Prillwitz et al., 2008), NGT (Ormel et al., 2010) o BSL (Cormier et al., 2012), entre otros. Pero en cualquier caso no hay nada comparable en tamaño y alcance a lo que existe para las lenguas vocales, no solo porque se trata de un campo de investigación con menor tradición, sino también por lo particularmente laboriosa que resulta la tarea de recopilar y anotar corpus de lengua de signos. Contamos desde hace varias décadas con estudios bastante detallados sobre diferentes

aspectos de la gramática de muchas lenguas de signos, pero son aún escasos o prácticamente inexistentes los recursos que proporcionan un análisis sintáctico sistemático y exhaustivo de muestras de uso real de alguna lengua de signos.

El corpus de lengua de signos australiana (Auslan), elaborado por Johnston y colaboradores (Johnston, 2010; Johnston y Schembri, 2006), se diseñó con el propósito general de estudiar la variación geográfica, y los objetivos se han extendido al análisis de muchos fenómenos gramaticales. Para estos propósitos se han transcrito las grabaciones utilizando un sistema de *id-glosas* (esto es, glosas que identifican unívocamente cada seña) y se anotan diferentes niveles de la estructura gramatical, con especial atención a la estructura argumental (Johnston, 2016). En concreto, se identifica el predicado y los argumentos para cada unidad que pueda ser analizada como cláusula. CLU (= ‘clause-like unit’) es la expresión utilizada para hacer referencia a estas unidades que pueden ser potencialmente cláusulas, pero cuya determinación es tentativa, puesto que se están tratando de determinar los patrones de la lengua a partir de los datos del corpus. Para cada CLU se especifica el número de los argumentos, su tipo y su posición dentro de la cláusula. También se ofrece información sobre la estructura de macroroles y roles semánticos de la cláusula. Gracias a la investigación de Hodge (2013), una parte del corpus Auslan se dotó de anotaciones para relaciones interclausales,¹ basadas en las propuestas de Halliday (2004).

En lo que concierne a la lengua de signos española, el proyecto RADIS (“Relaciones actanciales en discurso signado”), que sirve de base al análisis que aquí se presenta, está inspirado en el corpus Auslan y su anotación

¹Aunque la autora destaca el hecho de que se trata de una parte de la gramática donde los recursos codificados se combinan con el uso de procedimientos semióticos que las personas usuarias van usando para co-construir el discurso de una forma menos previsible en términos de patrones reconocidos, lo cierto es que el trabajo constituye una contribución importante en la identificación de dependencias interclausales.

(Cabeza et al., 2015, 2016). En el marco de este proyecto se han transcrito hasta ahora unas cuarenta grabaciones utilizando un sistema de id-glosas y dividiendo la transcripción en CLUs. La mitad de esas grabaciones contienen también anotación secundaria que incluye la categoría morfosintáctica de cada seña y la identificación de los argumentos de cada predicado.

El corpus de lengua de signos británica (BSL) está siendo actualmente anotado para su análisis sintáctico². Otros proyectos interesantes de anotación sintáctica de corpus de lenguas de signos se han desarrollado para de la lengua de signos de Finlandia (FinSL) (Jantunen et al., 2016), la lengua de signos polaca (PJM) (Rutkowski y Łozińska, 2016) y para la lengua de signos sueca (SSL) (Börstell et al., 2016). El corpus de esta última fue anotado inicialmente con la estructura argumental, de una manera análoga aunque no idéntica a la anotación de Auslan; pero en una segunda fase dio lugar a lo que por lo que conocemos es el primer intento de treebank de dependencias de una lengua de signos (Östling et al., 2017). La anotación básica se hizo con ELAN, con esquema de anotación propio; pero hay una versión con el modelo de UD. De hecho, la lengua de signos sueca es hasta ahora la única lengua de signos incluida en el repositorio de UD. Lo que se expone a continuación sobre sintaxis de dependencias de la lengua de signos española descansa sobre estos precedentes, y en particular sobre este último.

3. CORPUS ANOTADO, PROCEDIMIENTOS Y FORMATOS

La base que sustenta esta prueba piloto de anotación sintáctica es, como queda dicho más arriba, el proyecto RADIS, cuyo corpus está transcrito con id-glosas y parte de él está anotado con categorías morfosintácticas y aspectos básicos de la estructura argumental (Pérez, García-Miguel y Cabeza, 2019). De este corpus hemos seleccionado cuatro grabaciones con un total de 1400 *tokens* (id-glosas) para hacer pruebas con una anotación exhaustiva de relaciones sintácticas de dependencia.

² <https://bslcorpusproject.org>

Si definimos la dependencia como relación asimétrica entre dos palabras y asumimos que un árbol de dependencias tiene la propiedad de que un nodo puede tener varios nodos dependientes pero solo puede depender de un único nodo, lo único que necesitamos para anotar un corpus es un procedimiento que vincule cada palabra con aquella otra de la que depende (si es que depende de alguna, esto es, si no se trata del nodo raíz); y esto puede hacerse atribuyendo un identificador único a cada palabra. En la práctica las representaciones sintácticas suelen contener información adicional, en particular la categoría de cada elemento y el tipo específico de relación, esto es, la función sintáctica. Por eso, para codificar dependencias en cualquier lengua suele ser necesario como mínimo 4 campos (más el campo que contiene cada palabra/*token*/id-glosa), que deben convertirse en otras tantas columnas si organizamos el texto situando cada palabra en una línea:

- ID = identificador único de cada *token* / palabra
- *Token* = la palabra o la id-glosa que se está describiendo o relacionando
- CAT = categoría sintáctica de la palabra
- HEAD = núcleo/regente del que depende el *token* actual, expresado como identificador o como 0 (si se trata del *token* raíz)
- DEPREL = etiqueta con la relación específica entre dependiente y núcleo

Por ejemplo, la estructura sintáctica de la cláusula del corpus LSE recogida en (3) puede representarse gráficamente como en la Figura 1 y en columnas como en la Tabla 1.³

³Relaciones etiquetadas: *root* = raíz, palabra principal de la que dependen directa o indirectamente las demás; *nsubj* = sujeto nominal; *appos* = modificador aposicional; *obl* = complemento oblicuo (i.e., complemento verbal que no es ni sujeto ni objeto). La glosa del último ítem, *cl.e(2d)* hace referencia a una “clasificador de entidad” con la configuración de dos dedos doblados.

(3) “Un hombre se había subido a un árbol” [PS-JRV#005]

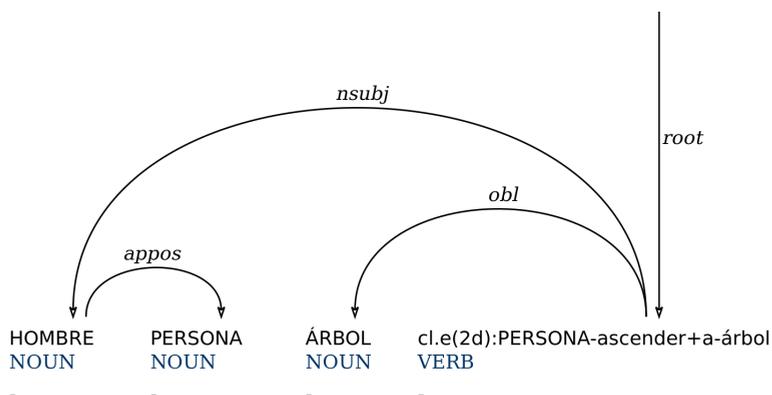
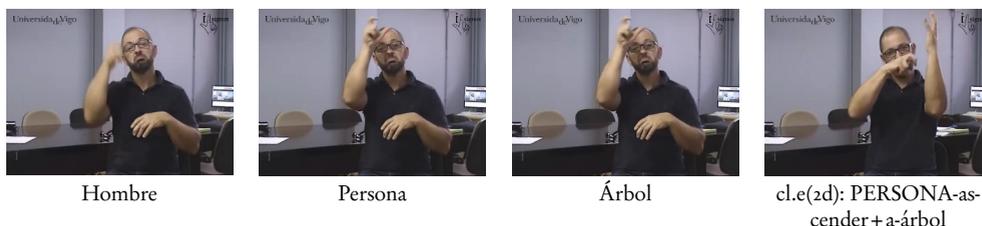


Figura 1: Diagrama de la estructura de dependencias del ejemplo de 3

ID	Token	CAT	HEAD	DEPREL
1	HOMBRE	Noun	4	nsubj
2	PERSONA	Nonun	1	appos
3	ÁRBOL	Noun	4	obl
4	cl.e(2d):PERSONA-ascender + a-árbol	Verb	0	root

Tabla 1. La estructura sintáctica del ejemplo (3) formateada en filas para cada *token* y columnas para categorías y relaciones.

Como el corpus está glosado utilizando el sistema de anotación multimedia ELAN, la anotación sintáctica puede hacerse inicialmente también en ELAN, añadiendo las líneas que sean necesarias para codificar relaciones de dependencia. El sistema de numeración automática de comentarios de ELAN permite asignar identificadores correlativos a cada *token* y en dos

líneas adicionales se incluirá el identificador del núcleo y la etiqueta de relación sintáctica.

	.000	00:00:11.000	00:00:12.000
Ref [127]	PS-JRV#005		
Trad_lit [127]	Un hombre estaba subido a un árbol (con ayuda de una escalera)		
MD_Glosa [316]	HO	PERSONA	ÁRBOL cl.e(2d):PERSONA-ascender+a-árb
MD_Cat [314]	N.P N.Loc	N	V.D
MD_id [315]	27 29	31	33
MD_Link [315]	33 27	33	0
MD_Rel [315]	nsu appos	obl	root

Figura 3. Anotación del ejemplo (3) en ELAN.

Para el intercambio de información con otros sistemas, y en particular si se pretende la integración con UD, hay que utilizar el formato CoNLL-U⁴ que organiza la información en 10 columnas y donde cada fila es un *token*. De los archivos anotados en ELAN se podrían derivar archivos en formato CoNLL-U. Alternativamente, puede hacerse primero la anotación con algún tipo de editor visual de dependencias y después importar desde ELAN. El formato CoNLL no plantea especiales problemas si, por ejemplo, anotamos solo relaciones entre id-glosas de la mano dominante. Pero sí plantea problemas que pueden ser irresolubles si se añaden relaciones con otros elementos simultáneos, sea la mano no dominante sea la expresión mediante componentes no manuales (expresión facial u otros movimientos corporales). En el apartado siguiente pasamos revista a algunos de esos problemas.

4. PROBLEMAS DE LA APLICACIÓN DE LAS SINTAXIS DE DEPENDENCIAS A LAS LENGUAS DE SIGNOS

4.1. Cuestiones generales

En líneas generales, asumimos que el tratamiento de las señas o signos como unidades léxicas (palabras en sentido lato) no debería suscitar de-

⁴ <https://universaldependencies.org/format.html>. El formato CoNLL-U es una versión revisada del formato CoNLL-X (Buchholz y Marsi, 2006).

masiados problemas teóricos ni prácticos para aplicar principios de sintaxis de dependencia a las lenguas de signos. Sin embargo, pronto aparecen dificultades en la identificación de las unidades sintácticamente relacionadas y en la definición de la naturaleza de su relación. Tienen que ver, por una parte, con características intrínsecas de las lenguas de signos y, por otra, con la todavía escasa tradición de estudios lingüísticos aplicados a su descripción, lo que se traduce en disquisiciones sobre la adecuación de determinadas etiquetas gramaticales.

4.2. *Problemas de tokenización*

Como hemos visto, en sintaxis de dependencias las relaciones se establecen entre palabras, concebidas como unidades elementales segmentables en la cadena hablada o escrita. Sin embargo, ha de tenerse en cuenta que la adaptación de soluciones de este tipo a una lengua de signos exige flexibilizar el concepto de segmento, puesto que en las diferentes líneas de anotación se superponen etiquetas atribuidas a elementos simultáneos, interpretados como morfemas. Por lo que se refiere a las características intrínsecas de las lenguas de signos, es preciso referirse a la simultaneidad de los medios de expresión, la iconicidad y el uso del espacio. Por razones prácticas, en las pruebas de anotación sintáctica de LSE, nos hemos limitado a las relaciones que se establecen entre señas segmentables, *tokens*, articuladas con la mano dominante, típicamente la mano derecha (MD_glosa), lo cual incluye también señas bimanuales. Se ha previsto la posibilidad teórica de analizar conexiones con señas articuladas en la mano izquierda (MI_glosa); pero es difícil gestionar un sistema con referencias y conexiones en líneas diferentes, y las dificultades aumentan si se quiere convertir la anotación a formato CoNLL o similar. Eso no impide que seamos conscientes de las limitaciones del procedimiento y de dónde están los principales problemas.

Debido a su producción gestual y su recepción visual, las lenguas de signos utilizan varios articuladores simultáneamente. Como vemos en (4),

puede ocurrir que ambas manos actúen una en relación a la otra, pero de tal forma que han de ser tenidas en cuenta para el análisis como unidades diferentes desde el punto de vista gramatical. Pero, además de las manos, otros articuladores están actuando al mismo tiempo, por lo que es necesario etiquetar también el movimiento de la mirada, las cejas, el torso o la cabeza.

(4) “(el niño) sale despedido y cae por un barranco” [HR-CBB#048]



Mano derecha= niño

Mano izquierda= barranco

Mirada de rol (sin contacto)

Boca abierta (acompaña el proceso verbal)

Todos estos elementos se anotan en líneas independientes en ELAN, donde para las señas de la mano no dominante (en general, la izquierda) que aportan contenido adicional también hemos previsto líneas con identificador de *token* (MI_id), conexión con un núcleo del que dependen (MI_link) y etiqueta de relación (MI_rel) siguiendo el modelo ilustrado para la mano dominante (derecha) en la figura 3. Sin embargo, el formato CoNLL presupone que las filas que corresponden a cada palabra están ordenadas secuencialmente, lo que no suele ser el caso en las lenguas de signos. Nuestro procedimiento de anotación de relaciones de dependencia en ELAN es similar al Swedish Sign Language Treebank, pero al convertir al formato CoNLL este asigna filas consecutivas tanto a señas de mano derecha como de mano izquierda, convirtiendo en secuenciales relaciones que son simultáneas. Como la solución no nos parece totalmente satisfactoria, de momento no la hemos implementado en nuestro *treebank* y nuestros archivos en formato CoNLL solo contienen señas articuladas con la mano derecha.

Por lo que se refiere al tratamiento de la iconicidad en las lenguas de signos, hay un aspecto concreto que ha suscitado un importante debate

terminológico. Se trata de representaciones como las del ejemplo (4), donde se seleccionan formas manuales que activan propiedades semánticas de los referentes y se combinan con parámetros de orientación, movimiento y componentes no manuales tales que, si bien se seleccionan de entre los que están disponibles en la lengua, dan como resultado construcciones de alta iconicidad. Las diferentes denominaciones que se utilizan revelan las maneras en que se conciben y la dificultad de destacar, entre las características definitorias de los elementos en cuestión, aquellas que son gramaticalmente pertinentes (Jarque, 2011): *verbos polimorfémicos* (Engberg-Pedersen, 1993) o *policomponenciales* (Slobin et al., 2003) llaman la atención sobre el aspecto morfológico, *clasificadores*, o *verbos / predicados clasificadores* (Supalla, 1986) inciden en su función semántica de categorizar entidades a partir de una configuración manual, en tanto que *verbos descriptivos o ilustrativos* ('depicting verbs', (Liddell, 2003) pone el acento sobre su potencial icónico de evocar referentes. El modelo semiológico que siguen Sallandre y Garcia (ver aportación en este mismo volumen) prefiere la denominación *unidades de transferencia*, que remite a su función de representar objetos de discurso y de asumir estrategias de cambio de rol a través de procedimientos altamente icónicos. Con independencia de la solución que se le dé al debate terminológico, que no es una cuestión menor, debe aclararse que estos elementos no son clasificadores de la frase nominal y aún está en discusión si comparten otras características de los clasificadores identificados como tales en lenguas como el mandarín. De momento pasan desapercibidos al análisis dependencial.

Otra cuestión relevante para la determinación de relaciones sintácticas está en relación con el uso gramatical del espacio. En concreto, una manera de expresar relaciones gramaticales consiste en modificar el movimiento o la orientación de un verbo (de la clase de los llamados direccionales o de concordancia). Esta capacidad del verbo de señalar un participante de la

acción proviene de la asociación entre una localización concreta en el espacio de signación y un referente (*locus* es el nombre que recibe esta asociación entre un lugar y un referente)⁵. Así, por ejemplo, un verbo como AYUDAR señala el agente y el beneficiario de la acción a través de los parámetros de la orientación y el movimiento. En este caso el debate tiene que ver con la naturaleza gramatical o gestual de esta relación. Algunos autores defienden que es una manifestación de concordancia, de tal manera que la articulación del verbo recibe una copia del locus referencial (o R-locus) vinculado a la frase nominal que hace las veces de agente o destinatario de la acción (Aronoff, Meir y Sandler, 2005; Lillo-Martin y Meier, 2011; Barberà Altimira, 2015). Otra aproximación (Rathmann y Mathur, 2002) prefiere explicar la concordancia a través de rasgos de persona y número que se codifican en la articulación manual. En cambio, la visión de Liddell (2000, 2003), muy influyente, defiende que estos verbos, llamados direccionales, señalan o indican hacia los referentes, tanto si estos están presentes en el espacio como si no (en este caso se asume que el espacio real es sustituido por otro espacio mental). Lo importante es que asistimos a un simple gesto de señalar en el espacio.

El problema de los verbos direccionales o de concordancia consiste pues en dilucidar si se debe tratar o no en el ámbito de la gramática. No obstante, la representación de las diferentes soluciones en el treebank es uno de los ítems que se pospone debido a que su expresión es de naturaleza simultánea.

4.3. Categorías léxicas

Una vez identificadas las unidades elementales, el primer paso para un análisis sintáctico, sea de constituyentes o de dependencias, sería la iden-

⁵El artículo de Martínez, Siyavoshi y Wilcox, en este mismo volumen, contiene una explicación, desde una perspectiva cognitiva, sobre la función que desarrolla la localización espacial en el establecimiento de una base nominal ('nominal grounding'), y en la posibilidad de retomar el mismo locus (o *Place*) en menciones subsiguientes.

tificación de la clase o categoría de cada unidad a partir de sus propiedades morfosintácticas. Sin embargo, en la LSE, como otras lenguas de signos, no es fácil establecer clases de palabras basándonos exclusivamente en criterios morfológicos. Un intento interesante de aplicación de criterios morfológicos a la categorización léxica en lenguas de signos ha sido propuesto por Schwager y Zeshan (2008), quienes utilizan el comportamiento de los verbos direccionales como eje de clasificación. No obstante, su análisis se combina con la aplicación (previa) de criterios semánticos y sintácticos. Los criterios puramente sintácticos tampoco resultan definitivos y encontramos en mayor medida que en las lenguas orales más próximas muchos signos polifuncionales, esto es, que pueden combinarse de diferentes maneras e interpretarse como nombres, verbos o adjetivos según el contexto (Pérez, García-Miguel y Cabeza, 2019). Por ejemplo, la seña identificada con la idglosa BICICLETA puede equivaler a “bicicleta” (Nombre) o a “andar en bicicleta” (Verbo), y la seña ROPA puede significar “ropa” (N) o “vestirse” (V). Sin embargo, entendemos que es posible asignar un valor categorial único a cada uso de un signo en el corpus. Para ello son imprescindibles criterios semánticos a partir de lo que nos parece una interpretación plausible de cada secuencia; pero también entran en juego indicios de muy diverso tipo: los *tokens* remiten siempre a una base de datos léxica donde está disponible la información categorial. Ocurre que, para un subconjunto de señas, esta es ambigua, y es preciso acudir a la realización real para analizar el *token*. A veces hay características morfológicas que ayudan a esa desambiguación (marcas morfológicas específicas de verbos o de nombres, por ejemplo), y en otras ocasiones es la construcción o, incluso, el significado general de la misma (incluidos los aspectos discursivos) la que permite decidir si se trata de una u otra categoría léxica. En la práctica esto quiere decir que el análisis sintáctico debe realizarse, al menos parcialmente, de manera simultánea a la desambiguación categorial. Pensando en aplicaciones computacionales, ca-

be la posibilidad de usar modelos probabilísticos de coocurrencia; pero para ello es necesario más corpus analizado, y nuestro objetivo es precisamente contribuir a esta necesidad.

En la Tabla 2 se ofrece la lista simplificada de categorías léxicas utilizada en la anotación gramatical del corpus. Como existe el propósito de ajustar la anotación sintáctica al modelo de UD, se ofrecen también equivalencias con las etiquetas categoriales de este sistema⁶.

Cat LSE	Descripción	UD_POS	Meaning
Adj	Adjetivo	ADJ	adjective
Adv	Adverbio	ADV	adverb
Aux	Auxiliar	AUX	auxiliary
Conj	Conjunción	CCONJ, SCONJ	(coord. / sub.) conjunction
Det	Determinante	DET	determiner
Loc	Locativo	ADV	adverb
MD	Marcador Discursivo	X	other
N	Nombre	NOUN	noun
Neg	Negación	PART	particle
Num	Número	NUM	numeral
Prep	Preposición	ADP	adposition
Posp	Posposición	ADP	adposition
Pro	Pronombre	PRON	pronoun
Pos	Posesivo	DET	determiner
V	Verbo	VERB	verb
Interr	(Pronombre) interrogativo	PRON	pronoun
Rel	(Pronombre) relativo	PRON	pronoun

Tabla 2. Categorías léxicas utilizadas en el proyecto RADIS (lista simplificada) y correspondencias con la lista de Universal Dependencies

En la Tabla 3 se ofrecen las frecuencias de cada categoría léxica en nuestra muestra, tomando como criterio su categorización con el sistema de UD

⁶“Universal POS tags”: <https://universaldependencies.org/u/pos/all.html>

UD_POS	Frecuencia
VERB	542
NOUN	465
ADV	86
ADJ	81
PRON	67
NUM	62
Otras (e indeterminadas)	99
Total	1402

Tabla 3. Frecuencia de las categorías léxicas principales

Las unidades de las lenguas de signos presentan algunas otras propiedades formales que pueden ser gramaticalmente relevantes y que se recogen como subcategorías en la anotación gramatical de la LSE, pero no se han recogido en la elaboración del treebank de dependencias. Es el caso de los verbos, para los que se anota el subtipo: descriptivo (V.D), direccional (V.Dir) o plano (V.P). También los nombres pueden ser descriptivos, localizables o planos (de Beuzeville, Johnston y Schembri, 2009). Estas distinciones morfológicas pueden afectar a las estrategias utilizadas para expresar relaciones sintácticas y, en menor medida, al rango de relaciones sintácticas posibles.

5. CONEXIONES Y RELACIONES DE DEPENDENCIA EN LSE

En el análisis y anotación de relaciones sintácticas de dependencia hay dos tipos de tareas: en primer lugar, determinar qué elementos están conectados y, en cada conexión, cuál es el regente y cuál es el dependiente; en segundo lugar, etiquetar el tipo específico de relación. Ambas tareas presentan aspectos problemáticos en su aplicación a la LSE. La primera tarea se refleja en los identificadores utilizados en el campo/línea HEAD/MD_link. La segunda tarea se refleja en el campo/línea DEPREL/MD_rel escogiendo de un listado cerrado de relaciones posibles. Las pruebas para el treebank

de LSE se están haciendo partiendo del listado de relaciones utilizadas en UD⁷, pero con algunas pequeñas adaptaciones a la lengua de signos. UD v.2 contiene una lista de 37 relaciones sintácticas universales que incluye relaciones funcionales de los argumentos nucleares (*nsubj*, *obj*, *iobj*, ...), dependientes no nucleares (*obl*, *advmod*, *aux*, *cop*, ...), dependientes nominales (*amod*, *nummod*, *det*, ...) y otras que no son relaciones de dependencia en el sentido más estricto (elementos coordinados [*conj*], *compound*, parataxis, ...). Una particularidad del sistema utilizado en UD es que en algunos casos diferencia relaciones sintácticas utilizando como criterio la categoría estructural del elemento dependiente, por ejemplo, sujeto nominal [*nsubj*] vs. sujeto clausal [*csbj*], o modificador nominal [*nmod*] vs. modificador adjetivo [*amod*] vs. modificador clausal [*acl*]. Este es un criterio que no compartimos del todo, pero que mantenemos para facilitar en el futuro la comparación interlingüística.

A continuación examinaremos algunas construcciones básicas para comentar relaciones cuya aplicación a la LSE resulta en algún aspecto problemática o al menos requiere alguna explicación.

5.1. *La estructura de la cláusula*

En general las señas que designan acciones o procesos pueden considerarse como equivalentes a verbos y tomarlas como nodo raíz de una estructura predicativa de la que dependen las señas que designan entidades que participan en ese evento y las que designan circunstancias del mismo. Con ese criterio es con el que hemos segmentado los textos en unidades similares a cláusulas (CLU – ‘clause like unit’). En esta prueba piloto no estamos teniendo en cuenta las relaciones interclausales, entre otras razones porque es difícil determinar si se trata de relaciones sintácticas o puramente discursivas. En el futuro podrán codificarse como conexiones de dependencia entre nodos raíz de CLU.

⁷ <https://universaldependencies.org/u/dep/all.html>

En este proceso de segmentación en CLUs e identificación del nodo raíz, surgen dos tipos de problemas:

- a Segmentos (CLUs) en los que no parece haber ninguna seña equivalente a un verbo
- b Segmentos (CLUs) en los que se encuentra más de un verbo

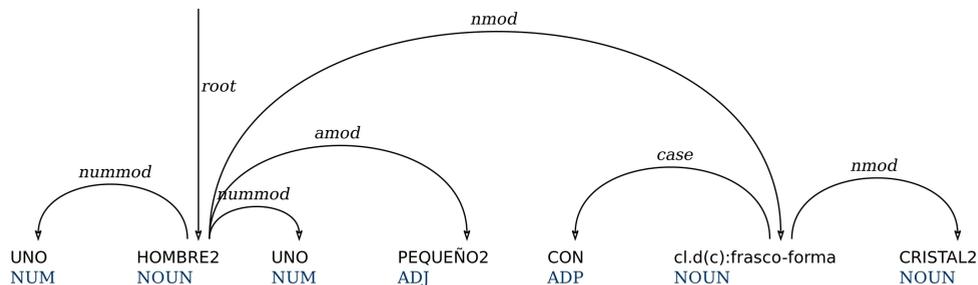
En la identificación del nodo raíz de segmentos identificados como CLU los casos problemáticos son, sobre todo, las predicaciones estativas y existenciales. En las predicaciones estativas encontramos como predicado (y por tanto, como nodo raíz de la estructura de dependencias) señas similares a nombres o adjetivos, que en otros contextos podemos encontrar como argumento de un predicado o como modificador de un nombre.

- (5) HOMBRE CANSADO
 Noun ADJ
 nsubj → root
 “El chico está cansado” [HR-AFS#008]

En las predicaciones existenciales y locativas podemos encontrar a veces una seña con el significado “haber” (6), pero también predicaciones sin verbo con una expresión referencial (7), que sirve de raíz para la estructura sintáctica, o predicaciones con una expresión referencial más una expresión locativa (8), donde tomamos la expresión locativa como predicado (‘root’) y la expresión referencial como sujeto.

- (6) *INDX HABER PERRO*
 advmod → root ← nsubj
 “Allí había un perro” [HC-JRV#048]

(7) “(Hay) un niño con un frasco de cristal” [HR-AFS#003]



(8) *TAMBIÉN* *AL-LADO* *PERRO*
 ADV ADV NOUN
 advmod → root ← nsubj
 “Al lado (estaba) su perro” [HR-ARS#003]

De un total de 462 segmentos “independientes” (CLUs) identificados, en 405 (88 %) el nodo raíz es un verbo, en 31 (7 %) un adjetivo, en 16 (4 %) un nombre o pronombre, y en 6 (1 %) otras clases de palabras.

Una segunda cuestión problemática es la posibilidad en LSE de que en una cláusula simple coexistan varios verbos en lo que claramente parece constituir una construcción de verbos en serie (Aikhenvald, 2018). Los verbos que aparecen en una serie verbal comparten valores de aspecto o polaridad (TAM en las versiones clásicas, difícilmente aplicables a las lenguas de signos por sus peculiaridades flexivas) pero la serie verbal es en sí misma una entidad de naturaleza sintáctica, aunque admite un cierto grado de gramaticalización de sus componentes.

(9) *MADRE* *AYUDAR* *HOMBRE* *PEQUEÑO2* *BAÑAR*
 V₁ V₂

“La madre baña al niño” [VV-FRE#073]

(10) *PERRO* *cl.c(B):patas-desplazarse* *NADAR*
 V₁ V₂

“El perro se acerca nadando” [HR-AFS#054]

En sintaxis de dependencias no existe una etiqueta estándar para el tipo de relación que se establece entre los verbos de una serie (distinta de la relación de auxiliar, y de cualquier relación predicado-argumento). En algunos treebanks de UD se utiliza un subtipo de la relación *compound*⁸ que se aplica a las series verbales (*compound:svc*); aunque existen diferencias importantes de comportamiento sintáctico entre las series verbales y otros compuestos.⁹ Hemos registrado 77 conexiones sintácticas etiquetadas como *compound:svc*.

Otras construcciones multiverbales son las construcciones de verbos en *sándwich*, en las que se repite un verbo antes y después de otro constituyente clausal. La conexión en este caso se ha etiquetado como *compound:vs* ('verb sandwich construction'). Finalmente, se registran también casos de simple repetición de un verbo. En (11) tenemos en una misma cláusula un ejemplo de cada tipo de construcción multiverbal.

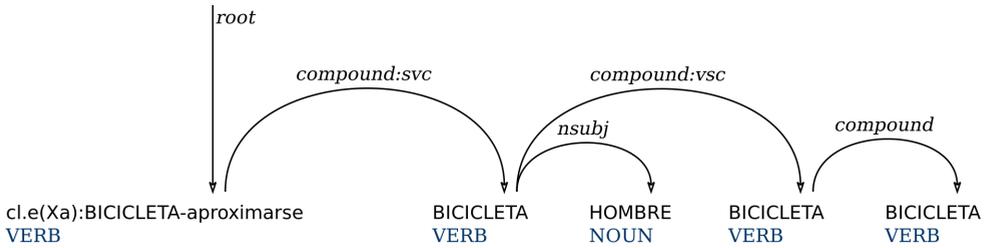
Los dos primeros verbos, uno descriptivo con clasificador de entidad y el otro un verbo simple, se combinan en una serie (*svc: serial verb construction*) para expresar conjuntamente la orientación y la manera del desplazamiento. El segundo se repite rodeando al sujeto en una construcción delimitativa en *sándwich* (*vs: verb sandwich construction*) característica de

⁸La relación *compound* se utiliza en UD para expresiones multipalabra similares a la frase inglesa *phone book*, si bien es preciso aclarar que hay otros dos tipos de expresiones multipalabra de formación más restringida, llamadas *fixed* ('as well as', 'plutôt que') y *flat* (nombres: 'Hillary Rodham Clinton', y fechas: '24 de diciembre').

⁹Los compuestos secuenciales son muy comunes en las lenguas de signos, y desde las primeras descripciones que se hicieron sobre la lengua de signos americana (ASL) por Klima y Bellugi (1979) se ha insistido en las adaptaciones formales que experimentan para asemejarse a las señas simples, de ahí que en el proyecto RADIS sean anotados como si fueran *tokens* simples. Consecuentemente, se incorporan a la base de datos léxica como signos de pleno derecho y no son objeto de un análisis sintáctico. Un ejemplo es el compuesto que se anota MENTE^NORMAL para mostrar los dos componentes de que procede. El primero, MENTE, es una seña monomanual que al entrar a formar parte del compuesto se acorta y reduce su movimiento. El significado no se deduce de la suma de las partes, pues es "inocente, ignorante".

muchas lenguas de signos (Fischer y Janis, 1990; Bø, 2010). La repetición final del verbo sirve en este caso para expresar iteración y continuidad de la acción.

(11) “El niño de la bici se iba acercando pedaleando a buen ritmo” [PS-JRV#028]

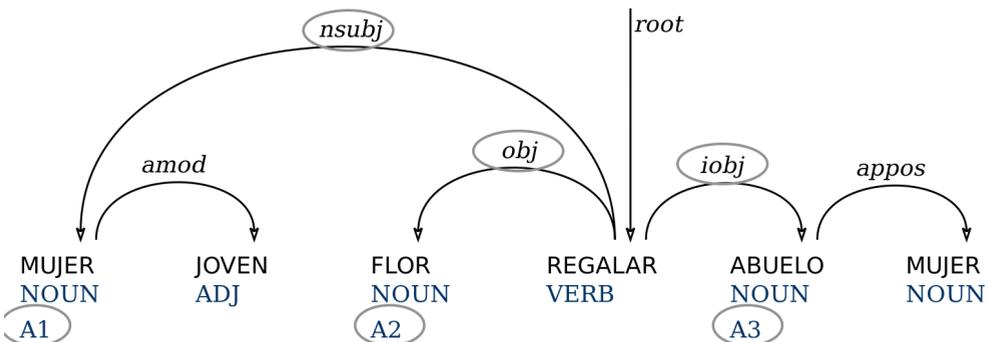


En la Tabla 4 se recogen las frecuencias registradas en los diferentes tipos de conexiones sintácticas que dan lugar a construcciones multiverbales.

DepRel		N
<i>compound:svc</i>	Construcción de verbos en serie	77
<i>compound:vsc</i>	Construcción de verbos en sándwich	17
<i>compound</i>	Repetición	10

Tabla 4. Relaciones sintácticas en predicados complejos.

(12) “La chica le regala una flor a su abuela” [VV-FRE#006]



Los elementos nominales que dependen de un predicado son sus argumentos, que pueden numerarse simplemente por orden de aparición, o etiquetarse con criterios semánticos; pero apenas existen propiedades formales que diferencien sistemáticamente en LSE las funciones sintácticas sujeto y objeto. En el proyecto RADIS se habían anotado previamente los argumentos de los predicados utilizando un sistema de numeración basado en criterios semánticos. Al anotar ahora relaciones de dependencia con el listado de UD estamos convirtiendo de manera sistemática los argumentos numerados (A_1 , A_2 , etc...) en relaciones etiquetadas sujeto, objeto, etc. Nos inclinamos por considerar como primer argumento (A_1), y etiquetarlo ahora como sujeto gramatical [nsubj], al argumento único de predicados monovalentes y, en el caso de predicados multivalentes, al argumento semánticamente más activo (“protoagente”) y como objeto [obj] al segundo argumento [A_2], semánticamente más pasivo o más afectado (“protopaciente”), con independencia el orden lineal. El tercer argumento [A_3], semánticamente receptor, de predicaciones de transferencia y comunicación, queda etiquetado como objeto indirecto [iobj].

Hay que tener en cuenta que en el discurso muchas veces un argumento queda implícito (o aludido por otros recursos expresivos), por lo que puede ser que en la secuencia de signos manuales el único argumento explícito sea el segundo, como en (13).

- (13) OLVIDAR SU SOMBRERO
 V A_2 [=obj]
 “Olvidó su sombrero” [PS-JRV#095]

Por otro lado, creemos necesario no tratar siempre al segundo argumento como Obj. Conviene distinguir entre Obj(eto) y Obl(ícuo), basándose en la diferente naturaleza de las entidades relacionadas (objetos o lugares) y no en la presencia o ausencia de un marcador relacional (adposición / caso) entre frase nominal y frase adverbial. Tratamos así como Obl todas

las expresiones que se refieren a lugares de referencia en construcciones descriptivas de desplazamiento y localización (Cabeza y García-Miguel, 2018), como la que encontramos en el ejemplo (3), repetido aquí como (15)

- (14) HOMBRE PERSONA ÁRBOL cl.e(2d):PERSONA-ascender+a-árbol
 A1 [=nsubj] A2 [=obl] V
 “Un hombre se había subido a un árbol” [PS-JRV#005]

Nótese que el verbo descriptivo incorpora en forma de clasificadores de entidad la entidad localizada, en la mano derecha, y el lugar de referencia, en la mano izquierda, esto es, la persona y el árbol respectivamente. Pero la anotación de relaciones sintácticas de dependencia no la estamos aplicando a la relación entre elementos simultáneos sino solo secuenciales, en este caso la relación del verbo con los nombres que explicitan sus argumentos.

La Tabla 5 resume la frecuencia de las relaciones sintácticas fundamentales en la cláusula y su orden relativo respecto al predicado.

DepRel		Dep > Nuc	Nuc > Dep	Total
<i>nsubj</i>	Sujeto nominal	205	25	230
<i>obj</i>	Objeto	49	50	99
<i>iobj</i>	Objeto Indirecto	6	6	12
<i>obl</i>	Oblicuo	23	13	36
<i>advmod</i>	Modificador Adverbial [=circunstancial]	51	29	80

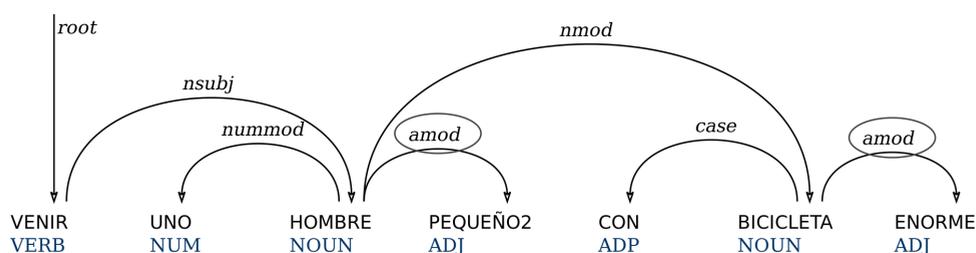
Tabla 5. Frecuencia de las relaciones sintácticas fundamentales en la cláusula y orden relativo del dependiente respecto del núcleo

Observamos que en este conjunto de datos la LSE muestra en el discurso un orden de palabras variable, donde solo el sujeto nominal muestra una clara preferencia por la posición preverbal, sin que ésta sea obligatoria. Los objetos ocurren casi con la misma frecuencia en posición preverbal o posverbal; mientras que los argumentos oblicuos y los modificadores adverbiales tienen cierta preferencia por la posición preverbal, aunque ocurren en más de un tercio de las ocasiones en posición posverbal.

5.2. Estructura de la frase nominal

Como argumentos de predicados tenemos normalmente nombres o grupos nominales, esto es, grupos de palabras que dependen directa o indirectamente de un nombre. Sin embargo, los estudios sobre las lenguas de signos se han ocupado poco en general por determinar qué elemento funciona como núcleo en un grupo nominal. Unas señas las categorizamos como nombres si ocurren frecuentemente aisladas como argumento de un predicado y otras señas las categorizamos como adjetivos si ocurren típicamente acompañando a nombres. Utilizamos aquí la etiqueta *amod* [modificador adjetivo], como en las conexiones de HOMBRE + PEQUEÑO y BICICLETA + ENORME de (15).

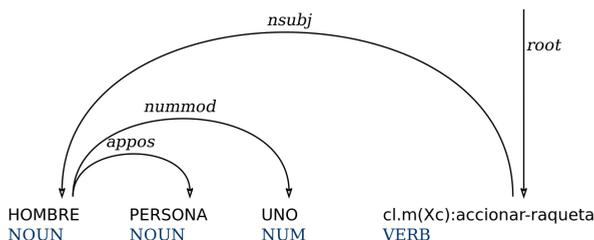
(15) “Viene un niño con una bicicleta enorme” [PS-JRV#026]



La combinación frecuente HOMBRE/MUJER + PEQUEÑO, con el significado de “niño/niña”, analizada como modificación adjetiva, es muy similar a la combinación también muy frecuente HOMBRE/MUJER + PERSONA. Aunque es discutible que se trate de construcciones diferentes, en este último caso entendemos que se trata de dos nombres, que se utilizan conjuntamente para referirse a una misma entidad por lo que diremos que están en aposición [*appos*]. En LSE es frecuente que en una frase nominal pueda combinarse un nombre plano con un nombre localizable. La primera de las dos palabras parece soportar mayor carga semántica; mientras que la segunda permite establecer un locus a partir del cual se establezcan

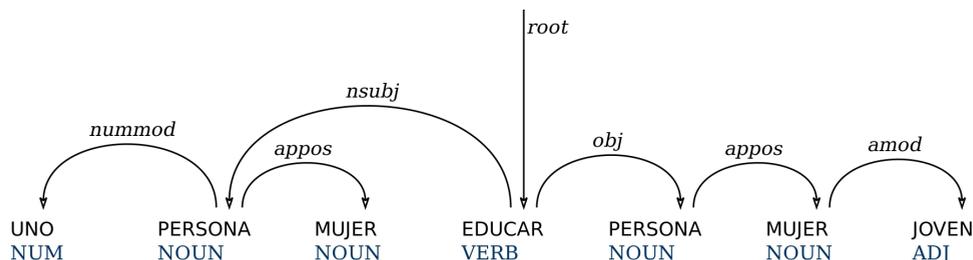
vínculos anafóricos. Tomaremos la primera como núcleo y la segunda como dependiente en aposición, a pesar de que esta suele tener mayor duración y parece tener mayor relevancia gramatical, al menos en el establecimiento de relaciones anafóricas y similares.

(16) “Un chico se entretiene con una raqueta” [PS-JRV#077]



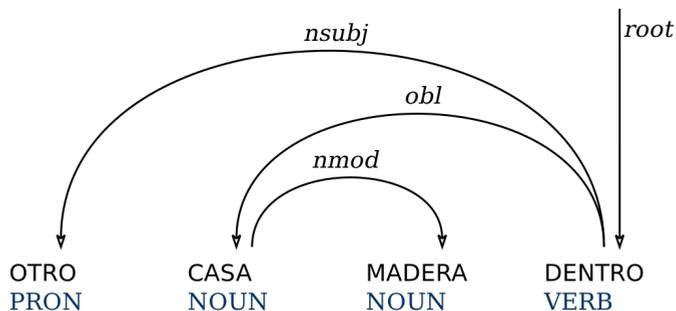
Ocasionalmente, registramos también PERSONA en primera posición y la analizamos como núcleo de la relación apositiva

(17) “Una mujer enseña a la chica” [VV-FRE#061]



Si los referentes son distintos y uno de ellos sirve para especificar al otro, estamos ante modificación nominal [*nmod*], como en la conexión de HOMBRE y BICICLETA en (15). En ese ejemplo el modificador nominal está introducido por la preposición CON; pero en muchas ocasiones se combinan nombres sin ningún tipo de conector, igual que en la modificación adjetiva y en la aposición. En la modificación nominal es el sentido el que determina la jerarquía de las relaciones y nos permite decidir qué nombre modifica a otro.

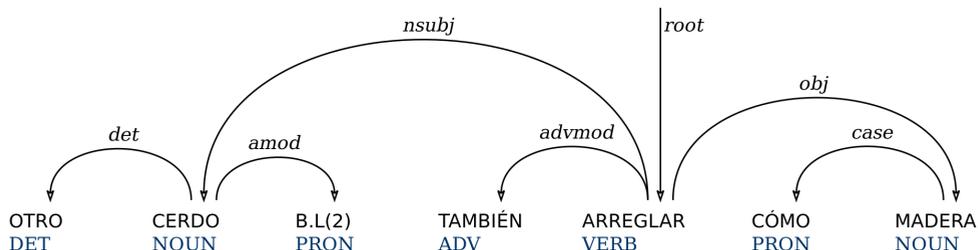
(18) “Otro entra en la casa de madera” [HC-JRV#053]



Otra cuestión problemática es el tratamiento de los elementos gramaticales que acompañan a nombres (determinantes, relacionantes, ...). En los trabajos sobre lenguas vocales hay quienes tienden a considerar los elementos gramaticales como núcleo de la construcción (como en el concepto de DP, habitual en los trabajos de sintaxis generativa) y quienes tienden a considerar estos elementos gramaticales como dependientes de un núcleo léxico (como se hace en UD). En nuestro análisis seguimos esta última vía y además de los diferentes tipos de modificadores, entendemos que de un núcleo nominal pueden depender determinantes (*det*), numerales (*nummod*) y relacionantes (*case*).

La relación *case* (= *case marking*) se usa en UD para “any case-marking element which is treated as a separate syntactic word (including prepositions, postpositions, and clitic case markers)”, como, por ejemplo, la preposición *of* en *the office of the Chair*, que es responsable de que la frase *the Chair* se interprete como un complemento nominal con significado de posesor con respecto al núcleo *office*. En LSE hay pocas adposiciones, como CON en (15), y muchas veces los nombres dependientes no necesitan ningún relacionante explícito, como vemos en el ejemplo (18). Pero en LSE, y también en otras lenguas de signos, las palabras interrogativas se usan frecuentemente como marcas para señalar un foco contrastivo, en ejemplos como el siguiente.

(19) “Otro cerdo, el segundo, también trabaja (en una casa de) madera” [HC-JRV#024]



En casos como este a la palabra interrogativa se le ha asignado la relación *case*, porque funciona a la manera de un introductor de un complemento. Se está haciendo no obstante una interpretación restrictiva de un procedimiento de pregunta-respuesta ampliamente utilizado en la sintaxis y en el discurso de las lenguas de signos (Jarque, 2016), de tal forma que únicamente se considera *case* la seña interrogativa en condiciones similares a las del ejemplo (19), en que se puede interpretar que relaciona un núcleo y un dependiente a la manera de un marcador de caso. Por otra parte, el empleo de la relación *case* se extiende también a otras señas en diferentes contextos, siempre como marcador de caso en la LSE.

La frecuencia de las diferentes relaciones sintácticas de los elementos dependientes de nombres aparece recogida en la Tabla 6. La alta frecuencia de los modificadores apositivos se explica por la alta frecuencia de las combinaciones con PERSONA, según hemos visto más arriba. En el caso de las aposiciones hemos tomado como núcleo siempre al primer elemento, aunque hemos visto que los mismos elementos léxicos admiten ordenaciones alternativas. En el resto de las conexiones, vemos que tampoco en las frases nominales existe un orden rígido, pero existe la tendencia clara de que los modificadores con contenido léxico (*amod*, *nmod*) sigan al núcleo, mientras que los elementos más gramaticales (*det*, *nummod*, *case*) tienden a preceder al núcleo nominal, aunque también se registren casos de lo contrario.

DepRel		Dep > Nuc	Nuc > Dep	N
<i>appos</i>	Modificador apositivo	0	82	82
<i>amod</i>	Modificador adjetivo	9	30	39
<i>nmod</i>	Modificador nominal	5	18	23
<i>nummod</i>	Numeral	28	12	40
<i>det</i>	Determinante	20	3	23
<i>case</i>	Relacionante	9	1	10

Tabla 6. Frecuencia de las relaciones sintácticas de dependientes de núcleos nominales y orden relativo del dependiente respecto del núcleo

6. CONCLUSIONES Y PERSPECTIVAS FUTURAS

En este artículo hemos presentado los primeros pasos para analizar sintácticamente un corpus de lengua de signos utilizando sintaxis de dependencias, hemos visto los conceptos y procedimientos básicos para la anotación, hemos examinado los principales problemas que presentan las lenguas de signos para el análisis y anotación de relaciones sintácticas, y hemos revisado, ejemplificado y recontado las relaciones sintácticas fundamentales en la estructura de la cláusula y de la frase nominal en un corpus de LSE.

Como experiencia piloto hemos anotado sintácticamente cuatro grabaciones que totalizan 1400 *tokens* y 462 CLUs. Este número es muy reducido si lo comparamos con los treebanks existentes para muchas lenguas con larga tradición oral y escrita; pero no es tan reducido si tenemos en cuenta las dificultades de recopilar para las lenguas de signos corpus anotado sintácticamente. El único treebank de una lengua de signos presente en el repositorio UD, el Swedish Sign Language treebank, cuenta con 1620 *tokens* y 203 *sentences*. En un futuro próximo nos proponemos ampliar el tamaño del corpus LSE anotado con relaciones de dependencia hasta un mínimo de 2000 *tokens*, revisarlo y adaptarlo escrupulosamente a las normas de UD, y subirlo al repositorio de UD para ponerlo a disposición de todos los lingüistas y personas interesadas.

Nuestra modesta contribución pretende aportar algún grano de arena a algo que consideramos aún muy necesario: disponer para lenguas de signos en general, y para la lengua de signos española en particular, de un conjunto creciente de muestras de uso glosadas, traducidas y analizadas. El conjunto de recursos que estamos recogiendo se caracteriza por el alto grado de detalle en la anotación, al asociar las grabaciones en vídeo con glosas identificativas de cada signo que permiten derivar del corpus un repertorio léxico de la LSE, con una traducción al español de cada fragmento que permite concebir el conjunto como un corpus paralelo y con un detallado análisis morfosintáctico y sintáctico que permite observar multitud de fenómenos característicos. Incluso si los análisis propuestos fueran discutibles, solo mediante su difusión y discusión podrá avanzarse en el conocimiento científico del léxico y la gramática de las lenguas de signos y de sus semejanzas y diferencias con otras lenguas.

Agradecimientos: Para esta investigación se ha contado con la financiación del Ministerio de Economía y Competitividad del Gobierno de España (FFI2013-41929-P; FFI2016-78194-P) y de la Consellería de Cultura, Educación y Ordenación Universitaria de la Xunta de Galicia (R2014/007; ED341D R2016/011; GPC2014/070; GPC2017/71).

REFERENCIAS BIBLIOGRÁFICAS

- AIKHENVALD, Alexandra Y. (2018), *Serial verbs*. Oxford: Oxford Univ. Press.
- ARONOFF, Mark, Irit MEIR y Wendy SANDLER (2005), "The paradox of sign language morphology". *Language*. 81/2, pp. 301-344.
- BARBERÀ ALTÍMIRA, Gemma (2015), *The meaning of space in sign language: reference, specificity and structure in Catalan Sign Language discourse*. Boston: De Gruyter Mouton.
- DE BEUZEVILLE, Louise, Trevor JOHNSTON y Adam SCHEMBRI (2009), "The use of space with indicating verbs in Auslan: A corpus-based investigation". *Sign Language & Linguistics*. 12/1, pp. 53-82. DOI: <https://doi.org/10.1075/sll.12.1.03deb>.
- BLOOMFIELD, Leonard (1933), *Language*. New York: Holt.
- BØ, Vibeke (2010), "Verb sandwich constructions in Norwegian sign language: a syntactic analysis". Master's Thesis, University of Oslo.
- BÖHMOVÁ, Alena, Jan HAJIČ, Eva HAJIČOVÁ y Barbora HLADKÁ (2003), "The Prague Dependency Treebank: A three-level annotation scenario", en A. Abeillé (ed.). *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Springer, pp. 103-127. DOI: https://doi.org/10.1007/978-94-010-0201-1_7.
- BÖRSTELL, Carl, Mats WIRÉN, Johanna MESCH y Moa GÄRDENFORS (2016), "Towards an Annotation of Syntactic Structure in the Swedish Sign Language Corpus", en E. Efthimiou et al. (eds.). *7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*. Paris: ELRA, pp. 19-24.
- BUCHHOLZ, Sabine y Erwin MARSÍ (2006), "CoNLL-X Shared Task on Multilingual Dependency Parsing", en *Proceedings of the Tenth Conference on Computational Natural Language Lear-*

- ning (*CoNLL-X '06*). Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 149-164.
- CABEZA, Carmen, José M. GARCÍA-MIGUEL, Carmen GARCÍA-MATEO y José Luis ALBA CASTRO (2016), "CORILSE: A Spanish Sign Language Repository for Linguistic Analysis", en *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portoroz, Eslovenia: ELRA, pp. 1402-1407.
- CABEZA, Carmen, José M. GARCÍA-MIGUEL, Juan R. VALIÑO, Cristina FREIRE y Ania PÉREZ (2015), "Relaciones Actanciales en Discurso Signado (RADIS)", en *Actas del Congreso CNLSE de la Lengua de Signos Española*. Madrid, pp. 64-77.
- CABEZA, Carmen y José M. GARCÍA-MIGUEL (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, pp. 227-263. DOI: <https://doi.org/10.7764/onomazein.41.13>.
- CHOMSKY, Noam (1957), *Syntactic Structures*. The Hague: Mouton.
- CHOMSKY, Noam (1965), *Aspects of the theory of syntax*. Cambridge (Mass.): MIT Press.
- CORMIER, Kearsy et al. (2012), "From Corpus to Lexical Database to Online Dictionary: Issues in Annotation of the BSL Corpus and the Development of BSL SignBank", en *Proceedings of LREC 2012*. Estambul, pp. 7-12.
- CROFT, William, D. NORDQUIST, K. LOONEY y M. REGAN (2017), "Linguistic Typology meets UniDependencies", en *Proceedings of the 15th International Workshop on Treebanks and Linguistic Theories (TLT15)*. Bloomington, pp. 63-75.
- ENGBERG-PEDERSEN, Elisabeth (1993), *Space in Danish Sign Language: The semantics and morphosyntax of the use of space in a visual language*. Hamburg: Signum.
- FISCHER, Susan y Wynn JANIS (1990), "Verb sandwiches in American Sign Language", en S. Prillwitz y T. Vollhaber (eds.). *Current trends in European sign language research*. Hamburg: Signum Press, pp. 279-294.
- GERDES, Kim, Bruno GUILLAUME, Sylvain KAHANE y Guy PERRIER (2018), "SUD or Surface-Syntactic UniDependencies: An annotation scheme near-isomorphic to UD", en *Proceedings of the Second Workshop on UniDependencies (UDW 2018)*. Brussels, Belgium: Association for Computational Linguistics, pp. 66-74. DOI: <https://doi.org/10.18653/v1/W18-6008>.
- HALLIDAY, M.A.K. (2004), *An introduction to functional grammar* 3rd ed. London: Arnold.
- HODGE, Grabielle (2013), "Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language)". Ph.D. thesis, Macquarie University.
- HUDSON, Richard A. (1984), *Word grammar*. Oxford: Basil Blackwell.
- JANTUNEN, Tommi, Outi PIPPURI, Tuija WAINIO, Anna PUPPONEN y Jorma LAAKSONEN (2016), "Annotated video corpus of FinSL with kinect and computer-vision data", en E. Eftimiou et al. (eds.). *7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*. Paris: ELRA, pp. 93-100.
- JARQUE, Maria Josep (2011), "Lengua y gesto en la modalidad lingüística signada". *Anuari de Filologia. Estudis de Lingüística*. 1, pp. 71-99. DOI: <https://doi.org/10.18653/v1/W18-6008>.
- JARQUE, Maria Josep (2016), "What about? Fictive question-answer pairs across signed languages", en E. Pascual Olivé y S. Sandler (eds.). *The conversation frame: forms and functions of fictive interaction*. Amsterdam: John Benjamins, pp. 171-192.
- JOHNSTON, Trevor (2010), "From archive to corpus: Transcription and annotation in the creation of signed language corpora". *International Journal of Corpus Linguistics*. 15/1, pp. 106-131. DOI: <https://doi.org/10.1075/ijcl.15.1.05joh>.
- JOHNSTON, Trevor (2016), "Auslan Corpus Annotation Guidelines". http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_February2016.pdf.
- JOHNSTON, Trevor y Adam SCHEMBRI (2006), "Issues in the creation of a digital archive of a signed language", en L. Barwick y N. Thieberger (eds.). *Sustainable Data from Digital Fieldwork: Proceedings of the Conference Held at the University of Sydney*. Sydney: Sydney University

- Press, pp. 7–16. KLIMA, Edward S. y Ursula BELLUGI (1979), *The signs of language*. Cambridge, Mass: Harvard University Press.
- LIDDELL, Scott K. (2000), “Indicating verbs and pronouns: Pointing away from agreement”, en K. Emmorey y H. Lane (eds.). *The Signs of Language Revisited*. Mahwah, N.J.: Lawrence Erlbaum Associates, pp. 303–320.
- LIDDELL, Scott K. (2003), *Grammar, gesture, and meaning in American Sign Language*. Cambridge; New York: Cambridge University Press.
- LILLO-MARTIN, Diane y Richard P. MEIER (2011), “On the linguistic status of ‘agreement’ in sign languages”. *Theoretical Linguistics*. 37/3-4, pp. 95–141. DOI: <https://doi.org/10.1515/thli.2011.009>.
- LONGACRE, Robert E. (1970), “Hierarchy in language”, en P. L. Garvin (ed.). *Method and theory in linguistics*. The Hague: Mouton.
- DE MARNEFFE, Marie-Catherine y Joakim NIVRE (2019), “Dependency Grammar”. *Annual Review of Linguistics*. 5, pp. 197–218.
- MEL’ČUK, Igor A. (1988), *Dependency syntax: theory and practice*. Albany: State University Press of New York.
- MEL’ČUK, Igor A. (2003), “Levels of dependency in linguistics description: Concepts and problems”, en V. Ágel et al. (eds.). *Dependenz und Valenz: ein internationales Handbuch der zeitgenössischen Forschung = Dependency and valency: an international handbook of contemporary research*. Berlin: Walter de Gruyter, pp. 188–229.
- NIVRE, Joakim (2006), *Inductive dependency parsing*. Dordrecht, The Netherlands: Springer.
- ORMEL, Ellen et al. (2010), “Glossing a multi-purpose sign language corpus”, en *Workshop Proceedings 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies.*, pp. 186–191.
- ÖSTLINGÇ, Robert, Carl BÖRSTELL, Moa GÄRDENFORS y Mats WIRÉN (2017), “Universal Dependencies for Swedish Sign Language”, en J. Tiedemann (ed.). *Proceedings of the 21st Nordic Conference on Computational Linguistics (NODALIDA 2017)*. Gothenburg: Linköping University Electronic Press, pp. 303–308.
- PÉREZ, Ania, José M. GARCÍA-MIGUEL y Carmen CABEZA (2019), “Anotación de corpus para o estudo da expresión gramatical de eventos: notas sobre o deseño do proxecto RADIS”. *Sensos-e*. 6/1, pp. 41–62. DOI: <https://doi.org/10.34630/sensos-e.v6i1.2563>.
- PRILLWITZ, Siegmund et al. (2008), “DGS Corpus Project – Development of a Corpus Based Electronic Dictionary German Sign Language / German”, en O. A. Crasborn et al. (eds.). *Proceedings of the 3rd Workshop on the Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora*. Marrakech, Morocco, pp. 159–164.
- RATHMANN, Christian y Gaurav MATHUR (2002), “Is verb agreement the same cross-modally?”, en R. P. Meier, K. Cormier y D. Quinto-Pozos (eds.). *Modality and structure in signed and spoken languages*. Cambridge: Cambridge University Press, pp. 370–404.
- ROJO, Guillermo y Tomás JIMÉNEZ JULIÁ (1989), *Fundamentos del análisis sintáctico funcional*. Universidad de Santiago de Compostela.
- RUTKOWSKI, Paweł y Sylwia ŁOZIŃSKA (2016), “Argument Linearization in a Three-Dimensional Grammar: A Typological Perspective on Word Order in Polish Sign Language (PJM)”. *Journal of UniLanguage*. 17/1, pp. 109–134.
- SCHWAGER, Waldemar y Ulrike ZESHAN (2008), “Word classes in sign languages: Criteria and classifications”. *Studies in Language*. 32/3, pp. 509–545. DOI: <https://doi.org/10.1075/sl.32.3.03sch>.
- SLOBIN, Dan I. et al. (2003), “A Cognitive/Functional Perspective on the Acquisition of ‘Classifier’”, en K. Emmorey (ed.). *Perspectives on classifier constructions in sign languages*. Mahwah, N.J.: Lawrence Erlbaum Associates, pp. 271–296.
- SUPALLA, Ted R. (1986), “The classifier system in American Sign Language”, en C. G. Craig (ed.). *Noun classes and categorization*. Amsterdam: John Benjamins, pp. 181–214.

TESNIÈRE, Lucien (1959), *Éléments de syntaxe structurale*. Paris: Klincksieck.
WELLS, Rulon S. (1947), "Immediate constituents". *Language*. 23, pp. 81-117.

recibido: septiembre de 2019
aceptado: noviembre de 2019