



Datos lingüísticos enlazados para facilitar la investigación en Lingüística: motivación, casos de uso y perspectivas

Julia Bosque-Gil

Universidad Politécnica de Madrid
{jbosque}@fi.upm.es

*Datos Enlazados y Multilingüismo para la Lingüística y las Humanidades
Digitales, 20.10.15, Escuela Técnica Superior de Ingenieros de
Telecomunicación, UPM, Madrid*

■ Corpus

- Escritos/orales, anotados o sin anotar, monolingües/multilingües, especializados/generales, sincrónicos/diacrónicos, etc.^[1]
- Ej. para español: CREA^[2], CORDE^[3], CORPES^[4], CEA^[5], CdE^[6], CORLEC^[7], CORELE^[8], COSER^[9] ...
→ Extracción de ejemplos, estadísticas de distribución de palabras o fenómenos, concordancias, estudios comparativos, etc.

■ Bases de datos

- Sintácticas, morfológicas, léxico-semánticas, tipológicas, prosódicas, fonéticas...
- Diccionarios: terminológicos, etimológicos, fundamentados teóricamente, etc.
- Ej. ADESSE^[10,11], AnCoraVerb^[12], AnCoraNom^[13], Morfogen^[14], AGLE^[15], DAELE^[16], DiCe^[17], Spanish FrameNet^[18], Atlas Interactivo de Entonación del Español^[19], Unified Verb Index^[20], UPSID-PC^[21], PDEV^[22], etc.
→ Consultas sobre determinadas características de una lengua, una unidad o una noción lingüística.

Problemas:

La mayoría de estos recursos...

1. no están vinculados
2. se desarrollan independientemente
3. no siempre resultan fácil de encontrar
4. están en distintos formatos
5. en ocasiones incluyen etiquetas diferentes para anotaciones similares
6. varían en la granularidad de la información

...pero pueden estar describiendo el mismo fenómeno o unidad lingüística.

Ejemplo. Verbo *hundir* en Adesse^[10]

<http://adesse.uvigo.es/data/verbos.php?sense=1989>

HUNDIR

Hacer que algo se vaya al fondo, deformar una superficie, derrumbar una construcción

->62 ejemplos<-

[-] Clasificación semántica y potencial valencial

Tipo de proceso:	Localización	Frecuencia
A0 INI	INICIADOR	31 (50 %)
A1 MOV	MÓVIL	62 (100 %)
A2 LOC	LUGAR	36 (58.1 %)
A3 BEN	Beneficiario	7 (11.3 %)

Perfil combinatorio >>

[-] Realizaciones valenciales (Esquemas sintáctico-semánticos):

Voz	Argumentos semánticos y Funciones sintácticas			N_ejemplos
HUNDIR _{act}	A0:INI = SUJ	A1:MOV = ODIR	A2:LOC = en LOC	13 >
HUNDIR _{act}	A0:INI = SUJ	A1:MOV = ODIR		7 >
HUNDIR _{act}	A0:INI = SUJ	A1:MOV = ODIR	A2:LOC = entre LOC	5 >
HUNDIR _{act}	A0:INI = SUJ	A1:MOV = ODIR	A3:BEN = OIND	3 >
HUNDIR _{act}	A0:INI = SUJ	A1:MOV = ODIR	A3:BEN = OIND	1 >
HUNDIRSE _{ref}	A0:INI = SUJ	A1:MOV = ODIR	A3:BEN = Refl	1 >
HUNDIRSE _{ref}	A0:INI = SUJ	A1:MOV = ODIR	A3:BEN = Refl	1 >
HUNDIRSE _{med}		A1:MOV = SUJ		17 >
HUNDIRSE _{med}		A1:MOV = SUJ	A2:LOC = en LOC	12 >
HUNDIRSE _{med}		A1:MOV = SUJ	A3:BEN = OIND	1 >
(ser) HUNDIDO _{pasiva}		A1:MOV = SUJ	A2:LOC = en LOC	1 >

[-] Ejemplos:

(solo 10, escogidos aleatoriamente, de un total de 62)

Verbo	Texto	Referencia	
HUNDIR	de nogal sin tapizar, con respaldo recto y brazos. Pero al viejo no le gusta el diván: en él se hunde . no hay firmeza, es para la blanda gente milanesa.	[SON:065.06]	+info
HUNDIR	, como muertas. De vez en cuando pasaba silenciosamente alguna bandada de grullas y tres cisnes hundían sus picos en ellas. Luego, la neblina fue en aumento y el lago	[CAR:014.22]	+info
HUNDIR	el acantilado sobre el mar. Y luego el agua turquesa, <i>transparente desde lo alto, en la que nos hundíamos</i> para emerger con algas que tú clasificabas, la calophyllis	[JOV:010.07]	+info

Ejemplo. Verbo *hundir* en AnCoraVerb_ES^[12]

AnCoraVerb_ES

Back | Go to lexicon letter a b c d e f g h i j k l m n ñ o p q r s t u v w x y z

hundir (verb, es) "hundir.lex.xml" in "AnCoraVerb_ES"

sense: 1 not lexicalized

verb.hundir.1.default lss: B21.unaccusative-state AncoraNet

PropBank	VerbNet	FrameNet	WordNet	Grouping
sink.01	other_cos-45.4	Not Aplicable	sink (2) sink (4)	sink.01

Arguments

Function	Argument	Theme	Constituents	sink.01(pb)	45.4(vn)
subj	arg1	tem		arg1	Patient
cc	argM	loc	sp(en)	argM	∅

Examples

- *0* se hunde en otras

sense: 2 not lexicalized

verb.hundir.2.default lss: A11.transitive-causative AncoraNet

PropBank	VerbNet	FrameNet	WordNet	Grouping
sink.01	other_cos-45.4	Not Aplicable	sink (2) sink (4)	sink.01

Arguments

Function	Argument	Theme	Constituents	sink.01(pb)	45.4(vn)
subj	arg0	cau		arg0	Agent
cd	arg1	tem		arg1	Patient
cc	argM	tmp	sp(para)	argM	∅

Examples

- hundir para siempre la piedra oscura

verb.hundir.2.anticausative lss: B21.unaccusative-state AncoraNet

Ejemplo. Verbo *hundir* en BabelNet^[23]



BabelNet

● Verb

[LOG IN](#) [REGISTER](#)

hundir|

SPANISH

TRANSLATE INTO...

SEARCH

[PREFERENCES](#)

Verb



caer, caerse, dejar caer, hundir

Caer o descender hacia un lugar o nivel más bajo.

ID: 00087367v | Concept



desmentir, hundir

Show (a theory or claim) to be baseless, or refute and make obsolete

ID: 00087875v | Concept



arruinar, destruir, estropear, hundir

Ruin utterly

ID: 00093632v | Concept



hundir, sumergir, sumir

Ejemplo. Relaciones morfológicas del verbo *hundir*^[24]:

<http://gedlc.ulpgc.es/investigacion/scogeme02/relmorfo.htm>



Resultado de las relaciones morfológicas

La palabra introducida **hundir** es infinitivo de la forma canónica **hundir** verbo transitivo pronominal.

Formas solicitadas sobre **hundir**:

[Grafo](#)

Relaciones sufijales

hundible: adjetivo -- obtenida con el sufijo *-ible*

hundición: sustantivo femenino -- obtenida con el sufijo *-ición*

hundido: adjetivo -- obtenida con el sufijo *-ido*

hundidor: sustantivo masculino -- obtenida con el sufijo *-idor*

hundimiento: sustantivo masculino -- obtenida con el sufijo *-imiento*

Relaciones prefijales

rehundir: verbo transitivo -- obtenida con el prefijo *re-*

[\[Flexión verbo \]](#) [\[Flexión sustantivo \]](#) [\[Flexión adjetivo \]](#) [\[Flexión otras formas \]](#) [\[Lematización \]](#) [\[Relaciones morfológicas \]](#)

[\[Desambiguador local \]](#) [\[Desambiguador morfosintático \]](#)

Enviar sugerencias a: [José R. Pérez Aguiar](#)

(C) [Grupo de Estructuras de Datos-UJPGC](#)

Ejemplo. Verbo *hundir* en el DiCe^[17]

causar que alguien sienta una ~ intensa (2 valores en total)

desesperación 1 *f.* (*Sentimiento*) [[ver ejemplos](#)] [[ver desesperación](#)]

hundir [a X en ART ~]

Glosa

causar que alguien sienta una ~ intensa

Ejemplos

1. Esto le hundió en la más profunda desesperación...La ira fue de tal intensidad que lanzó todos sus dardos venenosos al sexo contrario (web).

desvergüenza 1a *f.* (*Actitud*) [[ver ejemplos](#)] [[ver desvergüenza](#)]

causar que la ~ desaparezca (1 valor en total)

esperanza 1 *f.* (*Sentimiento*) [[ver ejemplos](#)] [[ver esperanza](#)]

hundir [ART ~]

Glosa

causar que la ~ desaparezca

Ejemplos

1. Georgia Burdyanadze pide ayuda al mundo para no hundir la esperanza de cambio.

Ejemplo. Verbo *hundir* en el DICTER^[25]

hundir

Variantes: *hundir*, *undir*.

(del lat. *fūdere* 'derramar', 'fundir'. Primeramente significó 'destruir, arruinar'; después se generalizó la acepción 'sumir, echar a fondo' por influjo de *hondo* (DECH)).

Familia léxica: confundir, confusamente, confusión, confuso, fundición, fundidor, fundiduría, fundir, hundición, hundido, hundidor, zahundir.

1. v. tr. u. t. c. prnl.

1ª datación del corpus: Ortega, *Conpusición Arismética y Geometría*, 1512.

Marca diatécnica: Metal.

Definición:

Derretir y licuar los metales, los minerales u otros cuerpos sólidos. (DRAE, s. v. *fundir*).

Sinónimos(s):

fundir₁.

Ejemplo(s):

Ejemplo 1:

Donde as de saber que cada marco de oro tenía de ley 20 quilates y cada marco de plata tenía de ley 5 dineros. Este mercader fizo **undir** todas tres piezas en una. (Ortega, *Conpusición Arismética y Geometría*, 1512, fol. 158r).

Ejemplo 2:

El hornillo ha de tener el suelo al medio de su alto y con algunos agujeros para que pueda la llama dar en la holla, porque se le da el fuego por abaxo, con retama y con hastillas muy mansamente y yqual, de manera que esté la holla recozida y no más, porque no **se hunda** el oro con mucho fuego. (Arphe, *Quilatador de la plata*, 1572, fol. 26v).

Ejemplo 3:

No tenían los indios peso alguno para pesar, [...]. Tampoco tenían moneda, teniendo mucho oro, plata, cobre, plomo, latón y estaño, aunque de aquestos tres postreros tienen poco, pero todo saben **hundir** y labrar. (Apiano, *Cosmographia*, 1575, fol. 75v).

2. v. tr.

1ª datación del corpus: Álaba, *Perfeto capitán*, 1590.

Marca diatécnica: Art.

Definición:

- Interoperabilidad semántica mediante el uso de ontologías y vocabularios lingüísticos
- Reutilización de los conjuntos de datos ya creados
- Integración de información lingüística...
 - ...sobre una o múltiples lenguas
 - ...dependiente o no de una teoría lingüística
 - ...perteneciente a distintos niveles de análisis lingüístico
 - ...originalmente en formatos diferentes
 - ...para facilitar su descubrimiento y acceso
- Mayor visibilidad
- Utilización en Procesamiento del Lenguaje Natural

- Repositorios: META-SHARE^[26], CLARIN^[27], ELRA^[28], FLaReNet^[29], LRE Map^[30], LingHub^[31], etc.
- Grupos de trabajo y proyectos: W3C Community Groups (Linked Data for Language Technology (LD4LT)^[32], Best Practices for Multilingual Linked Open Data (BPMLOD)^[33], W3C Ontology-Lexica^[34]), Open Linguistic Working Group (OWLWG)^[35], proyecto LIDER^[36], etc.
- Vocabularios: LexInfo^[36], OLiA^[37], GOLD^[38], etc.

Nube de datos lingüísticos enlazados: <http://linguistic-lod.org/lod-cloud>

- Consultas de toda la información sobre una lengua, una unidad o una noción lingüística disponible en recursos independientes
- Consultas con *restricciones* que combinen información de los distintos módulos lingüísticos (agregación de anotaciones sintácticas, semánticas, etc. provenientes de cada recurso)
..... para facilitar:
 - Los estudios contrastivos, tipológicos, de traducción, etc.
 - El análisis multiteórico
 - La búsqueda de paradigmas en Lingüística Teórica
 - Etc.

- Modelos y vocabularios para la representación de contenido de los distintos módulos lingüísticos, abordado desde distintas teorías y con distinta granularidad
- Interfaces que permitan al lingüista consultar los datos enlazados sin necesidad de aprender SPARQL
- Modelos y herramientas relativos al ciclo de vida de los datos: control de la calidad, mejora y extensión de los recursos por parte de la comunidad de lingüistas...



Muchas gracias

Referencias

- [1] Para tipos de corpus, consultar Llisterra 1999, Sinclair 1996.
- [2] REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. <<http://www.rae.es>>
- [3] REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. Corpus diacrónico del español. <<http://www.rae.es>>
- [4] REAL ACADEMIA ESPAÑOLA: Banco de datos (CORPES XXI) [en línea]. Corpus del Español del Siglo XXI (CORPES). <<http://www.rae.es>>
- [5] Carlos Subirats y Marc Ortega. 2012. Corpus del Español Actual <<http://sfncorpora.uab.es/CQPweb/cea/>>
- [6] Davies, Mark. (2002-) Corpus del Español: 100 million words, 1200s-1900s. Available online at <http://www.corpusdelespanol.org>.
- [7] CORLEC, Corpus Oral de Referencia del Español Contemporáneo, <http://www.llf.uam.es/ESP/Corlec.html>
- [8] Corpus Oral de Español como Lengua Extranjera (ELE), http://cartago.llf.uam.es/corele/home_es.html
- [9] Corpus Oral y Sonoro del Español Rural (COSER), <http://www.llf.uam.es/coser/index.php>
- [10] Base de datos de verbos, alternancias de diátesis y esquemas sintáctico semánticos del español (ADESSE), <http://adesse.uvigo.es/>
- [11] García-Miguel, José M. (2012): "Lingüística de corpus y valencia verbal", en I. Moskowich & B. Crespo (eds): Encoding the Past, Decoding the Future: Corpora in 21st Century. Newcastle: Cambridge Scholars Publishing, pp. 29-57.
- [12] AnCoraVerb_ES, http://clic.ub.edu/corpus/es/ancoraverb_es
- [13] AnCoraNom_ES, http://clic.ub.edu/corpus/es/ancoranom_es
- [14] Morfogen. Plataforma Web para el Estudio Morfogenético del Léxico. <http://bdme.usc.es/>
- [15] Archivo Gramatical de la Lengua Española de Salvador Fernández Ramírez (AGLE), <http://cvc.cervantes.es/lengua/agle/default.htm>
- [16] Diccionario de Aprendizaje del Español como Lengua Extranjera (DAELE), <http://www.iula.upf.edu/rec/daele/>
- [17] Diccionario de Colocaciones del Español (DiCe), <http://www.dicesp.com/paginas>
- [18] Spanish FrameNet (SFN), <http://spanishfn.org/>
- [19] Prieto, Pilar & Roseano, Paolo (coords). 2009-2013. Atlas interactivo de la entonación del español. <http://prosodia.upf.edu/atlasentonacion/>.
- [20] Unified Verb Index, University of Colorado, <http://verbs.colorado.edu/verb-index/>
- [21] The UCLA Phonological Segment Inventory Database (UPSID-PC), <http://www.linguistics.ucla.edu/faciliti/sales/software.htm>
- [22] Pattern Dictionary of English Verbs (PDEV), http://pdev.org.uk/#about_cpa
- [23] <http://babelnet.org/>
- [24] Aplicación para la obtención de relaciones morfológicas, Grupo de Estructuras de Datos y Lingüística Computacional del Departamento de Informática y Sistemas de la Universidad de Las Palmas de Gran Canaria, <http://gedlc.ulpgc.es/investigacion/scogeme02/reimorfo.htm>
- [25] DICTER: Diccionario de la Ciencia y de la Técnica del Renacimiento, Proyecto de Investigación de María Jesús Mancho Duque, Implementación realizada por el Dpto. de Sistemas de la Fundación General, <http://dicter.usal.es/>.
- [26] <http://www.meta-net.eu/meta-share>
- [27] <https://www.clarin.eu/content/repositories>
- [28] <http://catalog.elra.info/>
- [29] <http://www.flarenet.eu>
- [30] <http://www.resourcebook.eu/searchll.php>
- [31] linghub.lider-project.eu/
- [32] <http://www.w3.org/community/ld4lt/>
- [33] <http://www.w3.org/community/bpmlod/>
- [34] <https://www.w3.org/community/ontolex/>
- [35] linguistics.okfn.org/
- [36] www.lider-project.eu/
- [37] <http://www.lexinfo.net/ontology/2.0/lexinfo.owl>
- [38] <http://nachhalt.sfb632.uni-potsdam.de/owl/>
- [39] <http://linguistics-ontology.org/>



Datos lingüísticos enlazados para facilitar la investigación en Lingüística: motivación, casos de uso y perspectivas

Julia Bosque-Gil

Universidad Politécnica de Madrid
{jbosque}@fi.upm.es

*Datos Enlazados y Multilingüismo para la Lingüística y las Humanidades
Digitales, 20.10.15, Escuela Técnica Superior de Ingenieros de
Telecomunicación, UPM, Madrid*