

# Aplicación de recursos de xeración automática da lingua para estudos comparativos

## *Application of automatic language generation tools for comparative research*

María José Domínguez Vázquez  
Universidade de Santiago de Compostela / Instituto da Lingua Galega  
majo.dominguez@usc.es  
<https://orcid.org/0000-0002-6060-9577>

María Caiña Hurtado  
Universidade de Santiago de Compostela / Instituto da Lingua Galega  
maria.caina.hurtado@usc.gal  
<https://orcid.org/0000-0002-1224-6652>

Recibido: 18/01/2021

Aceptado: 05/06/2021

### Resumo

O principal obxectivo deste traballo é presentar unha das posibles aplicacións dos xeradores automáticos da lingua *Xera* e *XeraWord*: o seu emprego para a análise multicontrastiva. Estes prototipos plurilingües presentan a argumentación valencial de substantivos en español, francés e alemán —*Xera*— e en galego e portugués —*XeraWord*. No artigo achégase unha descrición da metodoloxía e da teoría na que se sustentan estes novos recursos valenciais lexicográficos e unha incursión no emprego dos recursos valenciais. A presentación de *XeraWord*, o máis recente dos prototipos, conxuga tanto a aproximación no manexo do recurso como os diferentes niveis de análise sintáctico-semántica que o conforman, os cales beben do simulador *Xera*. Co fin de exemplificar a aplicación comparativa botando man dos datos dos simuladores centrámonos na expresión da dirección no substantivo FUXIDA e os seus equivalentes en alemán, español e portugués. Esta análise serve para demostrar, por un lado, a pertinencia da metodoloxía empregada na construción dos xeradores e, por outro, a utilidade que estas ferramentas teñen á hora de levar a cabo estudos intralingüísticos, mais tamén inter e multilingüísticos.

### Palabras chave

Xeración automática da lingua, padróns argumentais, interface sintáctico-semántica, significado combinatorio, frase nominal

### Sumario

1. Introducción. 2. Notas lexicográficas sobre a argumentación sintáctico-semántica. 2.1. A necesidade de novos recursos valenciais. 2.2. Os dicionarios de

### Abstract

The main objective of this paper is to describe one of the possible applications of the automatic language generators *Xera* and *XeraWord*, that is, their use for multicontrastive analysis. These multilingual prototypes present the valency argumentation of nouns in Spanish, French and German —*Xera*— and in Galician and Portuguese —*XeraWord*. This article includes a description of the methodology and theory on which these new lexicographic valency resources are based, as well as an insight into the use of valency resources. The presentation of *XeraWord*, the most recent of the prototypes, includes both the approach on the use of the resource and the different levels of syntactic-semantic analysis that comprise it, which are based on the *Xera* simulator. In order to illustrate the comparative application using the data from the simulators, our focus lies on the expression of direction through the noun FUXIDA and its equivalents in German, Spanish and Portuguese. This analysis demonstrates, on the one hand, the relevance of the methodology used in the creation of the generators and, on the other hand, the usefulness of these tools when carrying out intralinguistic, but also inter- and multilingual studies.

### Key words

Automatic language generation, argument patterns, syntactic-semantic interface, combinatorial meaning, nominal phrase

### Contents

1. Introduction. 2. Lexicographical considerations on syntactic-semantic argumentation. 2.1. On the need for new valency resources. 2.2. Valency dictionaries as

valencias como dicionarios combinatorios. 3. Marco teórico e planos descritivos. 3.1. Descrición xeral de *XeraWord*. 3.2. A información sintáctico-semántica valencial nos xeradores. 3.2.1. Esquema argumental. 3.2.2. Escenas e roles semánticos. 3.2.3. O significado categorial e as clases semánticas. 4. Resultados sobre o emprego dos xeradores para o estudo lingüístico-comparativo. 4.1. Cuestións introdutorias. 4.2. Expresión da dirección de orixe. 4.3. Expresión da dirección de paso. 4.4. Expresión da dirección de destino. 5. Conclusións e perspectivas.

combinatory dictionaries. 3. Theoretical framework and descriptive levels. 3.1. General description of *XeraWord*. 3.2. Valency syntactic-semantic information in the generators. 3.2.1. Argument structure. 3.2.2. Semantic scenes and roles. 3.2.3. Ontological meaning and semantic classes. 4. Use of the generators for the linguistic-comparative study. 4.1. Introductory matters. 4.2. Expression of the direction of origin. 4.3. Expression of the direction of path. 4.4. Expression of the direction of destination. 5. Conclusions and further research.

## 1. Introducción

A literatura científica demanda máis e novos recursos con información de combinatoria argumental. As dificultades dun usuario ao enfrontarse a situacións de codificación da combinatoria léxico-argumental poñen de manifesto tamén esa necesidade (§2.1.). Nesta liña enmárcase o noso traballo, que ten como principal obxectivo presentar o simulador de xeración de padróns argumentais para o galego e o portugués —*XeraWord*— que achega non só unha nova aproximación onomasiolóxica á descrición da estrutura argumental nominal, senón que é novidoso desde un punto de vista metodolóxico.

Co obxectivo de amosar a aplicabilidade de *XeraWord* e *Xera* nos estudos contrastivos, compararemos a información recollida en *XeraWord* no que respecta á expresión da dirección no substantivo FUXIDA cos seus equivalentes nas outras linguas: o propio *XeraWord* achega datos tamén sobre o seu equivalente en portugués; mediante os datos que envorca o simulador *Xera* podemos completar o estudo comparativo coa inclusión no mesmo doutra lingua románica, o español, e unha lingua xermánica, o alemán (§4.). Para este fin partimos das escenas e roles semánticos como *tertium comparationis* multilingüe así como dos candidatos léxicos, os roles semánticos e clases semánticas prototípicas como información imprescindible para unha situación de codificación nunha lingua estranxeira, pero tamén como canles para describir o significado combinatorio lexemático-nominal (§3.).

Por tratarse dun prototipo recolle no seu estado actual un número reducido de unidades (§3.1.), se ben a automatización dos procedementos de procesamento e xeración da lingua —nos que se sustenta— contribuirá a un progresivo aumento dos substantivos analizados cara á súa implementación non só como recurso monolingüe para o galego e portugués, senón tamén como recurso comparativo entre estas dúas linguas e o alemán, francés e español, recollidas nos xeradores *Xera* e *Combinatoria* (Domínguez, Bardanca & Simões 2021, Domínguez 2020, Domínguez, Solla & Valcárcel 2019), impulsos imprescindibles para o deseño de *XeraWord*.

Tendo en conta o referido, nas seguintes seccións presentaremos unha breve aproximación teórica á argumentación sintáctico-semántica, centrándonos, especificamente, na descrición dos dicionarios de valencias e outros recursos valenciais (§2.). A seguir, describirase brevemente o marco teórico-descriptivo e a metodoloxía empregada na construción dos xeradores así como a interface de consulta de *XeraWord* (§3.). No capítulo (§4.) abórdase a análise contrastiva e as conclusións derivadas desta.

## 2. Notas lexicográficas sobre a argumentación sintáctico-semántica

### 2.1. A necesidade de novos recursos valenciais

Un escenario artéllase nunha rede de escenas (Fillmore 1977: 62), ás que están vinculados diversos conceptos e significados codificados mediante un conxunto de marcos sintácticos (Heringer 1984; Busse 2012). De tal xeito, na actualización dun concepto relativo á ‘Transferencia dunha posesión’ e, en concreto, nunha escena do tipo ‘alugar’ —fronte a ‘vender’ ou ‘comprar’— contamos con 4 entes centrais participantes: ‘Aquel que aluga’, ‘Aquilo que se aluga’, ‘A quen se lle aluga’ e ‘Durante canto tempo se aluga’. Esta escena diferénciase, por exemplo, dunha do tipo ‘vender’, na que non se contempla unha entidade temporal, mais si unha relativa ao custo da venda. Por tanto, dentro dun escenario do tipo ‘Transferencia dunha posesión’, contamos con diferentes escenas, que contemplan diferentes actantes e espazos conceptuais, que adoitan ser conceptualmente parellos nas linguas de culturas próximas, se ben poden diferir no plano formal e combinatorio nas diferentes linguas (Domínguez 2015). Esta aproximación está en consonancia co proposto por Smith (1991), que subliña que os falantes contamos con esquemas cognitivos comúns que contemplan diferentes tipos de situacións ou microescenas. Por tanto, nun acto comunicativo, un falante activa nunha determinada situación diferentes recursos expresivos, como, por exemplo, un lexema xunto co seu contorno sintáctico-semántico.

É precisamente esta relación de dependencia argumental —tanto no número como na tipoloxía— o obxecto de análise e descrición da Gramática e Lexicografía de Valencias, que parte da idea central de que un portador valencial conta cun potencial de espazos funcionais-valenciais que son actualizables de xeito obrigatorio ou facultativo (Engel 2004). Desta forma, a actualización concreta de cada unha das posibles lecturas dun lexema é o que determina o tipo e número de complementos argumentais así como as súas interaccións, os cales un aprendente de linguas estranxeiras ten que coñecer para garantir unha produción correcta desde un punto de vista sintáctico-semántico. Por exemplo, no alemán, ao contrario do que acontece no español ou no

galego, actualízanse diferentes verbos para expresar a acción de “roubar”: se o que se focaliza é o feito de que alguén roubou a alguén —roubo ao posuidor— úsanse *bestehlen* ou *ausrauben* (roubarlle con violencia todo a alguén, isto é, a totalidade da posesión que ten ou leva alguén nun determinado momento); mentres que no caso de que expresemos o obxecto ou posesión roubada cómpre usarmos *stehlen* ou *rauben* (que a diferenza de *stehlen* implica un proceso máis violento), entre outros. Por tanto, a diferenza do español ou do galego, no que un mesmo verbo contempla ámbalas dúas configuracións sintácticas —*roubar algo a alguén* ou *roubar alguén* (DRAG, s.v. roubar 1 e 2)—, o alemán recorre a lexemas verbais varios dependendo do esquema sintáctico-semántico e, en especial, do trazo ontolóxico actualizado, como, por exemplo, [obxecto] fronte a [humano]. Estes exemplos permítennos observar, por tanto, que na descrición do réxime e do potencial combinatorio non só a análise do plano sintáctico, senón tamén da valencia semántica —isto é, do significado combinatorio— é central (§3.2.).

Co fin de ilustrar estes padróns argumentais — desde unha perspectiva monolingüe (Busse 1994; Herbst, Heath, Roe & Götz 2004; VALBU 2004), contrastiva (VLdr 1983; VLdsp 2017) ou multilingüe (Portlex 2018)— naceron os dicionarios de valencias, estreitamente ligados ao ensino de linguas estranxeiras e á constatación de que as explicacións das obras gramaticais mediante un número reducido de exemplos non garanten que un falante dunha lingua estranxeira poida producir secuencias novas (Schumacher 2006). Por outra banda, os dicionarios estándar non achegan — nin é o seu principal cometido— información suficiente sobre a combinatoria léxica que permita a codificación (Bosque 2004; Pérez Serrano 2014) e aqueles orientados a falantes estranxeiros —DEPE (2002) ou DAELE— son nas linguas románicas aínda escasos. Así mesmo, non contamos para linguas como o galego e o portugués, por exemplo, con dicionarios ou ferramentas lexicográficas innovadoras que acometan o estudo das relacións valenciais e do potencial combinatorio, máis desenvolvidos, en especial, para a clase de palabra verbo no caso do alemán (VALBU 2004) e do inglés (Herbst et al. 2004), por exemplo<sup>1</sup>.

Non obstante, os substantivos e os adxectivos contan con escasos recursos lexicográficos actuais que describan a súa estrutura argumental. E todo isto pese a que, tal e como se subliñou previamente, no estudo-aprendizaxe de linguas estranxeiras cómpre adquirir o lexema xunto co seu contorno sintáctico-semántico (Laufer & Nation 2012), especialmente requirido nunha situación de produción. Precisamente,

---

1. Xunto cos dous dicionarios citados, existen ferramentas que describen relacións semántico-argumentais. É o inglés a lingua que conta con máis recursos deste tipo, como, por exemplo, *Verbnet* (Kipper, Dang & Palmer 2000), *Framenet* (Ruppenhofer et al. 2006) e *CPA* (Hanks 2004, Hanks/Pustejovsky 2005) para a descrición do verbo inglés, e *NomBank* (Meyers et al. 2004) para o nome tamén nesta lingua. Un exemplo para linguas románicas é *Ancora*: <http://clic.ub.edu/corpus/es/lexicones> [01/04/2021].



un nutrido número de erros na produción comunicativa de falantes estranxeiros débese a aspectos relativos á valencia e á combinatoria argumental (Sánchez & Jiménez 2013; Nied 2014; Gao & Liu 2020).

A literatura científica constata, pois, a necesidade de contar con recursos que describan a valencia e os padróns argumentais. Porén, estudos recentes sobre o manexo de ferramentas en rede en tarefas de corrección de erros motivados por interferencias lingüísticas (Müller-Spitzer et al. 2018; 2019) permítennos observar que para resolver cuestións relativas á valencia os estudantes universitarios —cun nivel A2-B1 en alemán como lingua estranxeira— non recorren a dicionarios de valencias<sup>2</sup>, senón a outro tipo de ferramentas en rede<sup>3</sup>. Así, nun 50 % das consultas os estudantes empregan dicionarios, mais á hora de resolver dúbidas sobre o réxime prefiren, tal e como amosa o Gráfico 1, programas de tradución —como *Google Translator* ou *Pons Translator*— e, en menor medida, recursos con textos paralelos, como *Linguee*<sup>4</sup>. Entre os dicionarios destacan Duden<sup>5</sup> e LEO<sup>6</sup>.

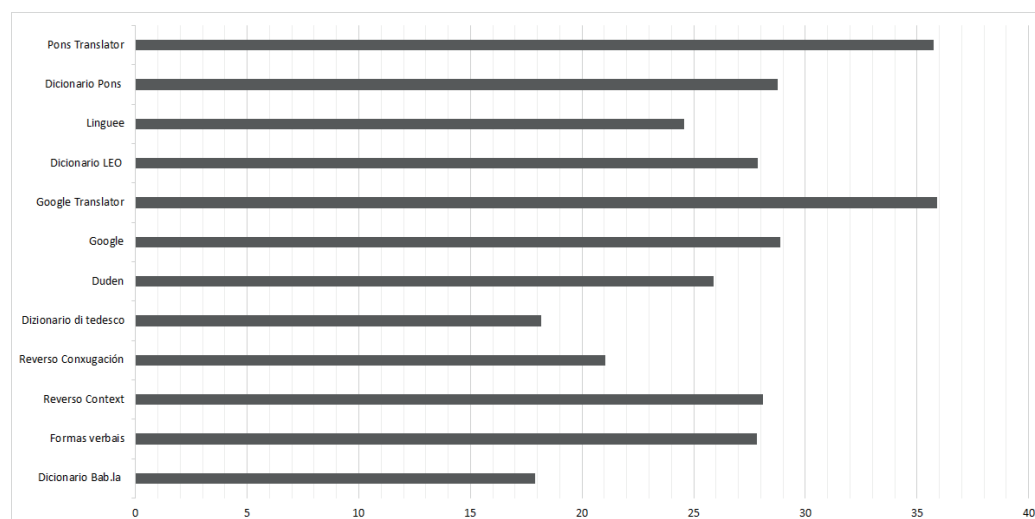


Gráfico 1. Recursos en rede empregados na consulta de dúbidas relativas ao réxime.

2. Cómpre subliñarmos que os recursos valencias en rede son escasos.

3. Estes datos resultan do proxecto *Recherchepraxis bei der Verbesserung von Interferenzfehlern aus dem Italienischen, Portugiesischen und Spanischen: Eine explorative Beobachtungsstudie mit DaF-Lernenden*. O estudo analiza os procedementos e recursos en rede manexados en tarefas de corrección de erros motivados por interferencias lingüísticas entre a lingua materna —lingua románica— e o alemán como lingua estranxeira. Os participantes foron estudantes universitarios da Università Roma Tre, da Universidade do Minho e da Universidade de Santiago de Compostela. Para máis información véxase Müller-Spitzer et al. (2019).

4. Dispoñible en: <https://www.linguee.es> [09/01/2021]

5. Dispoñible en: <https://www.duden.de> [09/01/2021]

6. Dispoñible en: <https://www.leo.org/alemán-español> [09/01/2021]

Do estudo tamén podemos tirar datos sobre o número de recursos manexados para resolver dúbidas relativas á valencia: no caso de resolución da interferencia (véxase nota a rodapé 3), os estudantes empregaron de media 5,8 ferramentas. Pola contra, cando os enquisados non conseguen solucionar o erro motivado pola interferencia na oración proposta para a súa corrección chegan a consultar uns 7,3 recursos en rede. En xeral, obsérvase que un 64,80 % das revisións propostas polos participantes no estudo resultaron incorrectas, mentres que só o 35,20 % foron correctas, isto é, solucionaron a interferencia causante da expresión errónea. Do estudo tamén resulta que o 42,5 % das consultas en ferramentas de tradución automática están relacionadas con unidades complexas ou multimembres (Müller-Spitzer et al. 2018: 300). Estes resultados non fan máis que constatar que en atención ao usuario son precisos novos recursos con información de combinatoria léxico-argumental.

## 2.2. Os dicionarios de valencias como dicionarios combinatorios

A descrición da valencia non se reduce a unha descrición cuantitativa —o número de espazos funcionais que pode abrir un portador valencial ao seu redor—, senón que os estudos valenciais tamén abordan cuestións tipolóxico-sintácticas e semánticas, así como a descrición do potencial combinatorio e diferentes aspectos distribucionais. En xeral, os estudos valenciais ocúpanse, por tanto, de determinar o catálogo de complementos e circunstantes así como a súa realización formal, os roles semánticos e os trazos ontolóxicos que poden aparecer cubrindo un determinado espazo funcional, esquemas argumentais e a interacción dos complementos, alén de restricións de combinatoria. A aplicación dos resultados da investigación valencial reflíctese nos dicionarios de valencias, que como obras lexicográficas combinatorio-sintagmáticas describen padróns argumentais —abstraccións lingüísticas de realizacións concretas. Por este motivo, mostran notables diferenzas de concepto con respecto a outros dicionarios orientados ao ensino de linguas estranxeiras —como DEPE (2002)— ou outros dicionarios de combinatoria —por exemplo, os de colocacións, como REDES (2004) ou DICE (2004).

A principal diferenza entre os dicionarios de combinatoria colocacional e valencial reside no concepto de combinatoria e nas teorías gramaticais concretas que os sustentan, e, en consecuencia, no caso dos dicionarios de valencia, na relación dependencial-valencial entre o predicado e os seus argumentos. Os dicionarios de valencias, en definitiva, describen argumentos específicos: aqueles requiridos por un tipo de clase ou subclase de verbo ou outro portador valencial, é dicir, os que permiten a expresión completa do predicado e contribúen á gramaticalidade e adecuación do comunicado. Como exemplo de dicionario de valencias do nome achegamos a

entrada do substantivo galego APRENDIZAXE (Táboa 1) do dicionario multilingüe da valencia do nome Portlex (2018), que en si mesmo xa constitúe un novo tipo de dicionario valencial en rede (Domínguez & Valcárcel 2020). APRENDIZAXE conta con tres argumentos específicos.

Aquel/Aquello que realiza una acción	
Realización formal: de	Frente a un modelo do pasado, centrado no ensino do profesor, explicador e ensinante de experiencias e de doutrinas, ofrécese unha perspectiva de variedade de métodos, de formas de facer, que anima máis a <b>aprendizaxe dos alumnos</b> .
Rasgo categorial: Humano, Animal	CORGA: López Herrerías, José Ángel e Tarrío Fernández, Xosé Antonio, <i>A trastenda pedagóxica do éxito escolar</i> Sada, Edicións do Castro, 1991
Tipo complemento: Complemento sujeto	
Aquel/Aquello no afectado	
Realización formal: de	Facíase tamén especial fincapé na <b>aprendizaxe do cálculo e da gramática española</b> , instrumentos considerados imprescindibles para triunfar no comercio e nos negocios que se suponía ían realiza-los alumnos unha vez cruzasen o océano.
Rasgo categorial: Inmaterial	CORGA: Sixirei Paredes, Carlos, <i>A emigración</i> , Vigo, Editorial Galaxia, 1988
Tipo complemento: Complemento objeto	
Clasificativo	
Realización formal: como	Se en troques mo publicasen nunha colección preciosa e me desen palmadiñas no lombo, seguro que a miña aprendizaxe <b>como poeta</b> sería moito peor porque pensaría que xa tiña todo feito.
Rasgo categorial: [-]: sin restriccións	CORGA: <i>A Nosa Terra</i> , ANT2007-02-08/46, Promocións Culturais Galegas S.A., 2007
Tipo complemento: Complemento nominal	

Táboa 1. Argumentos do substantivo gl. APRENDIZAXE no dicionario Portlex.

Cómpre subliñarmos que para a escolla dos argumentos que se presentan en cada entrada dos dicionarios de valencias non se aplica un criterio de frecuencia de coaparación<sup>7</sup> das pezas léxicas, como pode ser o caso noutros dicionarios de combinatoria, senón diferentes filtros sintáctico-semánticos, entre os que destaca a especificidade argumental. É precisamente o criterio de especificidade o que explica que, entre os 10 primeiros adxectivos, tanto en posición pre- como posnominal,

7. A non aplicación exclusiva deste criterio explica, por exemplo, que a pesar de que a expresión do complemento xenitivo no réxime verbal non é frecuente, é contemplada nos inventarios, logo que pode ser a única que permite o réxime do lexema verbal concreto.

obtidos do CORGA (§3.2.3.) para o substantivo APRENDIZAXE (Táboa 2) só nun exemplo o adxectivo desempeña a función de argumento ou complemento específico. Na práctica, isto supón que todos os demais adxectivos non se recollen nos nosos recursos como representantes dunha realización valencial na entrada concreta.

Lema adxectivo + Substantivo	Substantivo + Lema adxectivo
correcto <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> asociado <i>Adxectivo</i>
doado <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> continuo <i>Adxectivo</i>
duro <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> previo <i>Adxectivo</i>
lento <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> significativo <i>Adxectivo</i>
longo <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> permanente <i>Adxectivo</i>
bo <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> activo <i>Adxectivo</i>
adequado <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> intelectual <i>Adxectivo</i>
concienciado <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> infantil <i>Adxectivo</i>
constante <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> electrónico <i>Adxectivo</i>
continuo <i>Adxectivo</i> aprendizaxe <i>Substantivo</i>	aprendizaxe <i>Substantivo</i> social <i>Adxectivo</i>

Táboa 2. Selección dos 10 adxectivos máis frecuentes no CORGA en posición pre- e posnominal.

Na aplicación dos devanditos criterios radica unha diferenza importante fronte aos dicionarios de colocacións, que si que contarían nas súas descrições con adxectivos non específicos, como os recollidos na Táboa 2. Isto non significa, non obstante, que no noso modelo a clase adxectivo non se inclúa como posible realización de argumentos non específicos. Por tanto, contemplamos diferentes tipos semánticos de adxectivos actualizados como argumentos por diferentes clases nominais, como, por exemplo, *aprendizaxe humana* (a aprendizaxe dos humanos) ou *dor lumbar/estomacal* —onde *dor* escolle adxectivos que expresan partes do corpo<sup>8</sup>, fronte a, por exemplo, *dor crónica* ou *intensa* (Táboa 2). Por conseguinte, a frecuencia de coaparición dun lexema non é o criterio aplicado para contemplalo como representante dun argumento específico dun padrón argumental concreto, servindo, tal e como sinalamos, a especificidade de clase como criterio central.

En definitiva, no noso modelo descritivo —aplicado tanto no dicionario Portlex (2018) como nos xeradores *Xera* e *XeraWord* (§3)— o significado combinatorio e o significado categorial<sup>9</sup> son pezas centrais para a descrição do potencial combinatorio argumental.

8. A realización desta clase semántica argumental non está contemplada, por exemplo, no padrón argumental de APRENDIZAXE ou FUXIDA, de aí a súa especificidade en DOR.

9. O significado combinatorio —fronte ao inherente— comprende o significado relacional (roles semánticos) e o categorial (as categorías ontolóxicas que poden realizar os diferentes roles semánticos) (§3.2.2.).

### 3. Marco teórico e planos descritivos

#### 3.1. Descrición xeral de *XeraWord*

Os recursos de xeración automática de padróns argumentais da frase nominal en español, alemán e francés —*Xera* e *Combinatoria*—, nos que se sustenta *XeraWord*, naceron con dous firmes propósitos:

- Xerar padróns argumentais de xeito automático e, por conseguinte, exemplos que os reflecten para a súa incorporación no dicionario multilingüe Portlex (Domínguez & Valcárcel 2020) ou noutros recursos dixitais que teñan como obxectivo a descrición argumental.
- Servir de prototipos de novos tipos de dicionarios plurilingües de valencias, entendidos como recursos didácticos e ferramentas de información (Fuertes Olivera et al 2019: 79).

Ao non existir para o galego e o portugués recursos dixitais desta tipoloxía, o xerador piloto *XeraWord* (2020) baséase na metodoloxía e recursos ideados para os xeradores *Xera* e *Combinatoria* (Domínguez, Bardanca & Simões 2021, Domínguez 2020, Domínguez, Solla & Valcárcel 2019). Ademais, para o seu desenvolvemento foron deseñados *ad hoc* unha ferramenta de tradución do caudal léxico paradigmático —TraduWord— a partir dos datos extraídos de xeito automático do Wordnet e un flexionador nominal (Domínguez et al., no prelo) (§3.2.3.).

Se ben, no seu estado actual, *XeraWord* conta cun número moi reducido de entradas —5 substantivos en galego e 5 en portugués<sup>10</sup>—, a cantidade de información que achega cada lema analizado é significativa, de tal xeito que unha comparación multilingüe é factible (§4). Así, no caso do substantivo FUXIDA e os seus equivalentes en español, alemán e portugués, *Xera* e *XeraWord* contemplan os seguintes datos:

---

10. En concreto: gl. aumento | pt. aumento, gl. discusión | pt. discussão, gl. fuxida | pt. fuga, gl. olor | pt. cheiro e gl. presenza | pt. presença. A súa escolla xustifícase polo seu papel de representantes de diferentes escenas cognitivas ou campos semánticos. Especificamente, a escolla do substantivo *olor* para o galego, en lugar do equivalente *cheiro*, responde ao carácter marcado que pode ter este último (empregado como sinónimo de *fedor*) en determinados contextos e/ou variedades xeográficas da lingua. Isto non implica, non obstante, que a ferramenta non poida enriquecerse no futuro con esta e outras formas sinónimas dos lemas inicialmente escollidos.

	Padróns sintáctico-argumentais simples <sup>11</sup>	Estructuras sintáctico-semánticas <sup>12</sup>	Candidatos léxicos: lemas <sup>13</sup>	Candidatos léxicos: formas <sup>14</sup>
<b>esp. HUIDA</b>	17	82	2319	10979
<b>dt. FLUCHT</b>	27	111	1928	24971 <sup>15</sup>
<b>gal. FUXIDA</b>	18	88	1868	10547
<b>pt. FUGA</b>	13	59	1480	5287

Táboa 3. Datos achegados polos xeradores para fuxida e os seus equivalentes.

*XeraWord* resulta o máis novidoso pola súa accesibilidade onomasiolóxico-conceptual aos datos, dado que presenta a categorización ontolóxica dos argumentos e a súa realización formal. A consulta de datos realízase mediante un menú despregable, no que, unha vez seleccionada a lingua e o substantivo, podemos obter información sobre as diferentes realizacións argumentais dunha categoría ontolóxica, como, por exemplo, {humano}, {lugar} ou {evento}. No caso de FUXIDA e para a categoría {lugar} obtemos as seguintes subcategorías que se poden observar na parte inferior da Figura 1.



Figura 1. Menú despregable na interface de *XeraWord*.

11. Un exemplo é: determinante + adxectivo o + huida en singular + adxectivo o + a + actante N4. Cada padrón argumental vai acompañado en *Xera* dun exemplo estándar dunha das clases semánticas que poden aparecer representando o actante N4, como, *por exemplo*:

*lugar poboación - la {inesperada} huida {inesperada} a la capital*

12. Recóllense aquí datos sobre a interface sintáctico-semántica (realización formal e clase semántica), como, por exemplo: o N4 de FUXIDA aparece realizado coa clase semántica {lugar poboación}.

13. Trátase da adxudicación dos lemas ás clases semánticas, por tanto, *fillo* é un lema da categoría semántica {animado humano familia}.

14. É a atribución de formas ás clases semánticas, por tanto, *fillo, fillos, filla, fillas* son formas de representantes léxicos da categoría semántica {animado humano familia}.

15. O alemán presenta un número maior de formas léxicas por mor da declinación.



Realizada a escolla, por exemplo, {poboación}, o recurso amosa as diferentes realizacións e subcategorizacións semánticas (Figura 2).

Idioma

Núcleo

Filtro semántico

Tipo de lugar

Límite de frases para xerar: 250

	Anotación semántica	Exemplo
<input type="checkbox"/>	país nome propio a {desesperada} fuxida {desesperada} de España	país nome propio a {frenética} fuxida {frenética} a Cuba
<input type="checkbox"/>	cidade nome propio a {desesperada} fuxida {desesperada} por Ancara	cidade nome propio a {frenética} fuxida {frenética} a Cuba
<input type="checkbox"/>	cidade nome propio a {desexada} fuxida {desexada} de Estocolmo	cidade nome propio a {inesperada} fuxida {inesperada} a Madrid
<input type="checkbox"/>	país nome propio a {rápida} fuxida {rápida} por Andorra	país nome propio a {desesperada} fuxida {desesperada} ao Brasil
<input type="checkbox"/>	país nome propio a {recente} fuxida {recente} do Congo	poboación a {sospeitosa} fuxida {sospeitosa} á cidade
<input type="checkbox"/>	urbanismo ponte a {frenética} fuxida {frenética} pola ponte	cidade nome propio a {inminente} fuxida {inminente} ata Madrid
<input type="checkbox"/>	país nome propio a {desesperada} fuxida {desesperada} polo Brasil	país nome propio a {frenética} fuxida {frenética} ata Cuba
<input type="checkbox"/>	poboación a {curiosa} fuxida {curiosa} da aldea	país nome propio a {desexada} fuxida {desexada} ata o Canadá
<input type="checkbox"/>	cidade nome propio a {desesperada} fuxida {desesperada} desde Ancara	poboación a {repentina} fuxida {repentina} ata a capital
<input type="checkbox"/>	país nome propio a {triste} fuxida {triste} desde Chile	cidade nome propio a {inesperada} fuxida {inesperada} cara a Madrid
<input type="checkbox"/>	país nome propio a {perigosa} fuxida {perigosa} desde a China	país nome propio a {frenética} fuxida {frenética} cara a Cuba
<input type="checkbox"/>	poboación a {irremediable} fuxida {irremediable} desde a capital	

Figura 2. Realizacións e categorías semánticas para gl. FUXIDA para {lugar poboación} en *XeraWord*.

Seleccionando, por exemplo, a categoría {país nome propio} (*a {frenética} fuxida {frenética} a Cuba*), e premendo en *Xerar frases*, a ferramenta envorca datos de combinatoria simple argumental, exportables en formato CSV e JSON (Figura 3).



Xerar Frases
A fuxida a Etiopía
A fuxida a Tailandia
A fuxida a Suecia
A fuxida a Arabia Saudita
A fuxida a Bulgaria
A fuxida a Albania
A fuxida a Estonia
A fuxida a Suacilandia
A fuxida a Eslovenia
A fuxida a México
A fuxida a Qatar

Exportar CSV    Exportar JSON

Figura 3. Exemplo de envorcado de datos en *Xera Word*.

Como vemos, o manexo da ferramenta resulta bastante amigable para o usuario, que pode acceder aos datos dun xeito rápido e sinxelo. Porén, para obter unha visión máis ampla do funcionamento do xerador, cómpre termos en conta tamén os niveis lingüístico-descriptivos nos que se fundamentan os xeradores. Co fin de facer explícita esta información, preséntanse a continuación unha serie de conceptos-chave.

## 3.2. A información sintáctico-semántica valencial nos xeradores

### 3.2.1. Esquema argumental

Desde un punto de vista formal, as realizacións dos argumentos específicos dependen dos diferentes recursos expresivos dos que dispoñan as linguas. Por exemplo, no caso de FUXIDA contamos cos seguintes esquemas argumentais asociados aos catro roles semánticos do seu padrón<sup>16</sup> (Táboa 4).

16. A descrición segue a corrente valencial aplicada en Portlex (Domínguez & Valcárcel 2020). Abreviaturas: N=Nome, A=Adxectivo, DET=Determinante. Actualmente os xeradores envorcan datos para a posición DET tendo en conta o artigo determinado. Os adxectivos da tipoloxía {Adxectivo} son non específico-argumentais e resultan de datos de frecuencia dos corpus, sen incluírse os posesivos (véxase Cidrás 2017).

<b>Argumento 1:</b> <b>Aquel/Aquilo</b> <b>que realiza a</b> <b>acción</b>	N1	DET	{Adxectivo}	núcleo	{Adxectivo}	de	DET	N1
		A	posible	fuxida	posible		do	exército
	A1	DE	{Adxectivo}	núcleo	{Adxectivo}	de		N1
		A	rápida	fuxida	rápida	de		Pedro
	N2	DET	{Adxectivo}	núcleo	{Adxectivo}	de	DET	N2
		A	sospeitosa	fuxida	sospeitosa		do	país
<b>Argumento 2:</b> <b>Locación: orixe</b>	N2	DET	{Adxectivo}	núcleo	{Adxectivo}	de		N2
		A	frenética	fuxida	frenética	de		Francia
	N2	DET	{Adxectivo}	núcleo	{Adxectivo}	desde	DET	N2
		A	nova	fuxida	nova	desde	a	cidade
	N2	DET	{Adxectivo}	núcleo	{Adxectivo}	de		N2
		A	repentina	fuxida	repentina	desde		Portugal
<b>Argumento 3:</b> <b>Locación: paso</b>	N3	DET	{Adxectivo}	núcleo	{Adxectivo}	por	DET	N3
		A	frenética	fuxida	frenética		polo	mar
	A3	DET	{Adxectivo}	núcleo	{Adxectivo}	por		N3
		A	perigosa	fuxida	perigosa	por		Dublín
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	a	DET	N4
		A	desesperada	fuxida	desesperada		á	cidade
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	a		N4
		A	perigosa	fuxida	perigosa	a		París
<b>Argumento 4:</b> <b>Locación:</b> <b>destino</b>	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	cara a	DET	N4
		A	desexada	fuxida	desexada		cara ao	monte
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	cara a		N4
		A	repentina	fuxida	repentina	cara a		Berlín
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	para	DET	N4
		A	rápida	fuxida	rápida	para	o	Brasil
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	para		N4
		A	frenética	fuxida	frenética	para		Portugal
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	ata	DET	N4
		A	rápida	fuxida	rápida	ata	o	Brasil
	N4	DET	{Adxectivo}	núcleo	{Adxectivo}	ata		N4
		A	perigosa	fuxida	perigosa	ata		Portugal

Táboa 4. Esquemas argumentais de gl. FUXIDA.

Dado que a estrutura argumental é unha moeda de dúas caras, non só a descrición da realización formal dos argumentos, senón tamén a dos roles semánticos é esencial (3.2.2.): trátase de elementos bisagra que permiten, por unha banda, adxudicar o material expresivo a unidades abstractas de orde xerarquicamente superior, como son as escenas e o escenario; por outra banda, posibilitan o procedemento á inversa, isto é, pasar da abstracción á descrición do material expresivo. Unha análise completa do significado combinatorio precisa tamén incluír na descrición o significado categorial (3.2.3.).

### 3.2.2. Escenas e roles semánticos

Un escenario cobre diferentes recursos expresivos dun concepto (§2.1.) e, consecuentemente, expresións supralexemáticas ou lexemáticas de diferente índole<sup>17</sup>. Nun nivel abstracto, unha escena (Domínguez 2015) é un conxunto de elementos que recollen o coñecemento relevante para a comprensión e produción de enunciados lingüísticos e significados expresados mediante formas aínda por definir (cf. Hermanns & Holly 2007: 109; Busse 2009: 84). Nun nivel inferior de abstracción, as escenas sintáctico-semánticas —compostas por unha forma concreta e un significado concreto— son estruturas mínimas expresivas que evocan diferentes escenas cognitivas (Heringer 1984; Ziem 2008), en concreto predicados (nominais) —como exemplo FUXIDA—, e, por tanto, situacións (cfr. Smith 1991). Para a aproximación contrastiva, o escenario e as escenas actúan a modo de *tertium comparationis*, permitindo analizar unha rede de construcións lingüísticas compatibles desde un punto de vista conceptual e semántico.

Seguindo o catálogo de roles proposto por Engel (2004) e a súa aplicación multi-contrastiva (Domínguez & Valcárcel 2020) contemplamos 4 roles centrais —Axentivo, Afectado, Clasificativo e Locativo—, cuxa indexación<sup>18</sup> permite, ademais, describir de xeito máis pormenorizado o plano formal<sup>19</sup>. Son os roles semánticos os elementos que permiten adxudicar o material expresivo aos escenarios e escenas, dado que estas se sustentan na presenza dun rol focalizado<sup>20</sup> e, por tanto, esta relación non é arbitraria (Engel 1996: 227). Nesta liña, para determinar a existencia dunha escena de ‘Desprazamento’ é precisa a presenza implícita ou explícita dun rol semántico locativo-directivo —independentemente da súa realización formal. Este elemento é o que garante non só a análise da estrutura semántico argumental relacional (Táboa 4), senón tamén o estudo interlingüístico.

---

17. De tal xeito, no modelo descritivo xeral (Domínguez 2015) un concepto como ‘Movemento’ pode ser expresado por diferentes estruturas (*Fóra!*, *Marcha fóra!*, *O rapaz marchou para América* etc.) e pode denotar diferentes tipos de movemento —desprazamento, modo de desprazamento, movemento nun lugar fixo, entre outros (véxase tamén Talmy 2020 e a súa coñecida clasificación *satellite-framed languages* fronte as *verb-framed languages*).

18. Para exemplificar esta indexación tomamos o rol Locativo: fronte ao locativo situacional, o locativo direccional desglósase en orixe, paso e destino/dirección. Para o catálogo de roles véxase Portlex (2018).

19. A diferenza do dicionario Portlex (2018), os xeradores automáticos non presentan explicitamente os roles semánticos, aínda que estes forman parte do aparato teórico-descriptivo conformando o significado combinatorio (§2.2.).

20. Por tanto, a clase semántica xeral do predicado, non é o elemento chave para a adxudicación do material expresivo aos escenarios e escenas, tal e como amosan os seguintes usos perifrásticos, que non forman parte dun escenario de “Desprazamento”: “Pepita anda triste ultimamente” ou “Meteuse a enfermeira”.

### 3.2.3. O significado categorial e as clases semánticas

Son as propiedades do predicado as que determinan o paradigma de candidatos léxicos que poden cubrir un determinado espazo valencial-funcional para a expresión dun rol determinado. Porén, as descrições categoriais nos dicionarios de valencias —e non unicamente— resultan excesivamente xerais e, por tanto, poden ser imprecisas desde un punto de vista descritivo-lingüístico, así como de pouca axuda —tamén por imprecisión— nun proceso de produción nunha lingua estranxeira. Así, por exemplo, se ben a categoría ontolóxica {animal} co substantivo APRENDIZAXE é axeitada, grazas ao noso coñecemento enciclopédico contemplamos que os cans aprenden, pero menos prototipicamente os percebes, e, tamén, que os percebes non poden fuxir dun lugar, a diferenza das aves. Tamén sabemos que hai certas combinatorias léxicas argumentais máis prototípicas, isto é, frecuentes: CON APRENDIZAXE asociamos a aprendizaxe dun estudante, por exemplo, ou dunha temática ou área de coñecemento (a aprendizaxe das matemáticas).

Así pois, dado que o predicado selecciona os seus argumentos e o conxunto de pezas léxicas compatibles, tanto para a xeración automática dos datos como para o manexo dos recursos, é necesario describir e organizar este caudal léxico ontolóxico e relacionalmente, así como establecer clases semánticas. Por este motivo, os xeradores contemplan dous niveis semánticos descritivos do material léxico que ocupa os diferentes actantes valenciais: os prototipos léxicos e as clases semánticas.

Co fin de coñecer os prototipos léxicos —e de aí extraer as clases semánticas que adoitan aparecer nos diferentes argumentos dun nome (§3.1.)—, nos xeradores resultou esencial partirmos de datos lingüísticos da lingua real. No caso de *Xera Word*, para a extracción dos datos lingüísticos e de frecuencia do galego empregamos o *Corpus de Referencia do Galego Actual* (CORGA)<sup>21</sup>. Por exemplo: para unha estrutura como a exemplificada na frase *o aumento do prezo* (isto é: *DET + aumento + de + DET + SUBS*) precisamos listaxes de todas as posibles unidades léxicas que poderían aparecer na posición do SUBS, ordenadas pola súa frecuencia de uso ou ocorrencia no corpus<sup>22</sup>.

---

21. Trátase dun corpus documental aberto que ten como obxectivo achegar datos da lingua real actual e que contén textos (escritos e orais) datados entre o 1975 ata a actualidade (Domínguez, López & Barcala 2019: 179). O corpus está etiquetado automaticamente e contén na actualidade máis de 40 millóns de palabras. A ferramenta de consulta permite realizar diferentes tipos de buscas e extracción da información, tendo en conta as formas, os lemas ou os elementos gramaticais. Dispoñible no enderezo: <http://corpus.cirp.gal/corga/> [09/01/2021].

22. A escolla deste corpus para a extracción de datos do galego responde ao feito de que o CORGA, polas súas funcionalidades, permite extraer este tipo de información de xeito automático a través das consultas por *expresións coincidentes*. Este tipo de consulta permite especificar unha estrutura concreta (por exemplo: *lema=aumento + de + DET + SUBS*) e obter resultados que teñan en conta a frecuencia dos lemas contidos na cadea buscada, sen importar contexto concreto no que se inserte a frase. Polo

Deste xeito, a ferramenta devolve xa resultados ordenados por frecuencia, como se pode ver na imaxe a seguir:

		Coincidencias	Documentos	Frec. norm.
1	<b>aumento</b> Substantivo de Preposición o Artigo <b>prezo</b> Substantivo	55 / 48.184.012	46 / 52.650	1/millón
2	<b>aumento</b> Substantivo de Preposición o Artigo <b>temperatura</b> Substantivo	49 / 48.184.012	42 / 52.650	1/millón
3	<b>aumento</b> Substantivo de Preposición o Artigo <b>número</b> Substantivo	47 / 48.184.012	43 / 52.650	1/millón
4	<b>aumento</b> Substantivo de Preposición o Artigo <b>paro</b> Substantivo	43 / 48.184.012	37 / 52.650	1/millón
5	<b>aumento</b> Substantivo de Preposición o Artigo <b>nivel</b> Substantivo	35 / 48.184.012	32 / 52.650	1/millón
6	<b>aumento</b> Substantivo de Preposición o Artigo <b>demanda</b> Substantivo	32 / 48.184.012	29 / 52.650	1/millón
7	<b>aumento</b> Substantivo de Preposición o Artigo <b>poboación</b> Substantivo	26 / 48.184.012	22 / 52.650	1/millón
8	<b>aumento</b> Substantivo de Preposición o Artigo <b>gasto</b> Substantivo	25 / 48.184.012	20 / 52.650	1/millón
9	<b>aumento</b> Substantivo de Preposición o Artigo <b>consumo</b> Substantivo	21 / 48.184.012	17 / 52.650	0/millón
10	<b>aumento</b> Substantivo de Preposición o Artigo <b>custo</b> Substantivo	19 / 48.184.012	18 / 52.650	0/millón
11	<b>aumento</b> Substantivo de Preposición o Artigo <b>renda</b> Substantivo	17 / 48.184.012	9 / 52.650	0/millón
12	<b>aumento</b> Substantivo de Preposición o Artigo <b>productividade</b> Substantivo	16 / 48.184.012	11 / 52.650	0/millón
13	<b>aumento</b> Substantivo de Preposición o Artigo <b>producción</b> Substantivo	16 / 48.184.012	9 / 52.650	0/millón
14	<b>aumento</b> Substantivo de Preposición o Artigo <b>taxa</b> Substantivo	15 / 48.184.012	14 / 52.650	0/millón
15	<b>aumento</b> Substantivo de Preposición o Artigo <b>emprego</b> Substantivo	14 / 48.184.012	13 / 52.650	0/millón
16	<b>aumento</b> Substantivo de Preposición o Artigo <b>desemprego</b> Substantivo	14 / 48.184.012	14 / 52.650	0/millón
17	<b>aumento</b> Substantivo de Preposición o Artigo <b>capacidade</b> Substantivo	13 / 48.184.012	13 / 52.650	0/millón
18	<b>aumento</b> Substantivo de Preposición o Artigo <b>volume</b> Substantivo	13 / 48.184.012	12 / 52.650	0/millón
19	<b>aumento</b> Substantivo de Preposición o Artigo <b>presión</b> Substantivo	12 / 48.184.012	10 / 52.650	0/millón
20	<b>aumento</b> Substantivo de Preposición o Artigo <b>imposto</b> Substantivo	12 / 48.184.012	10 / 52.650	0/millón

Figura 4. Exemplo de busca por *expresións coincidentes* no CORGA.

Este tipo de resultados que devolve o CORGA son similares aos que se obtiveron para as outras linguas empregando Sketch Engine<sup>23</sup> (Kilgarriff et al. 2014), que tamén conta cunha funcionalidade similar<sup>24</sup>. Deste xeito, o proceso de extracción dos datos lingüísticos é practicamente equivalente ao das outras linguas de traballo. Se comparamos a figura 4 coa 5, veremos que tanto o CORGA (para o galego) como o Portuguese Web 2011 (prTenTen11) do Sketch Engine permiten extraer automaticamente a información de frecuencias dun xeito similar:

tanto, a pesar de que as características dos datos que ofrece o prTenTen11 (constituído por unha grande cantidade de textos procedentes da web) e o CORGA (constituído por un número moito menor de documentos orais e escritos, contemplando unha maior tipoloxía de textos, desde xornalísticos até literarios) son diferentes, as similitudes nas súas funcionalidades de procura e filtrado da información resultan útiles para os datos que precisabamos obter.

23. <https://www.sketchengine.eu> [01/01/2021]

24. Sketch Engine, polo momento, non inclúe un corpus etiquetado para o galego, polo que non foi posible empregar esta ferramenta para o conxunto de linguas analizado.



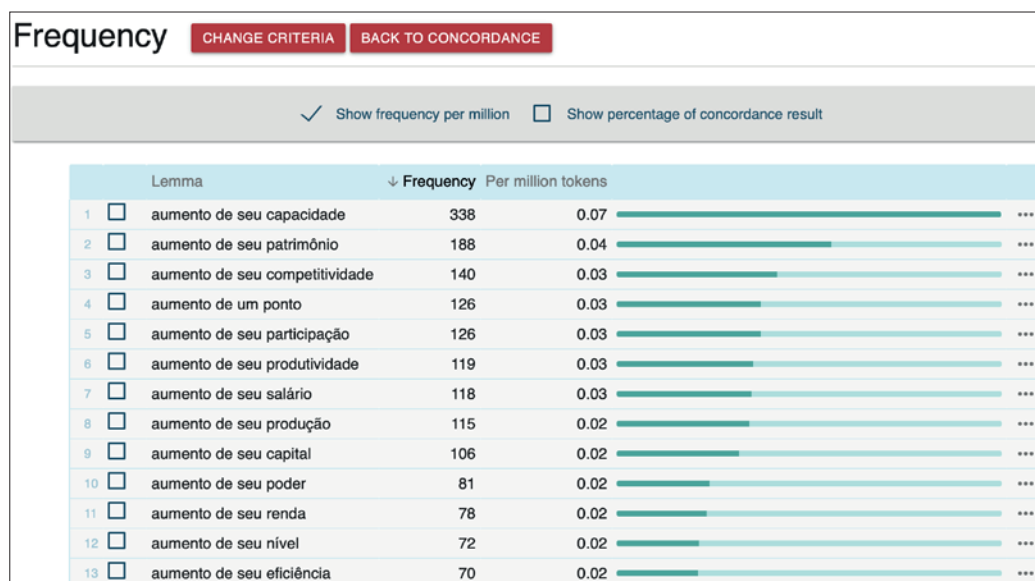


Figura 5. Exemplo de busca por frecuencias no prTenTen11 de Sketch Engine.

Partindo destes datos de frecuencia e dun cribado semántico-valencial con atención á actualización concreta do rol focalizado polo substantivo nunha acepción determinada (§2.2.), lévase a cabo un procedemento de etiquetaxe ontolóxica a partir do cal podemos establecer as denominadas clases semánticas (Domínguez, Solla & Valcárcel 2019), que resultan da suma dos trazos categoriais dos candidatos léxicos<sup>25</sup>. Por tanto, describen de modo agrupado o vocabulario que, en relación paradigmática, expresa os diferentes actantes. Por exemplo, os datos recollidos relativos a “Fuxida + dirección a onde” (Argumento 4 na Táboa 5) permiten establecer clases semánticas prototípicas, tais como:

25. Trátase, por tanto, dunha ontoloxía *bottom up* realizada manualmente a partir do fichado de datos de corpus. Para propostas sobre taxonomías léxicas plurilingües véxase a ferramenta *Kind* (<http://www.tecling.com/cgi-bin/kind/2020/index.pl?lang=spanish> [15/01/2021]) ou novas propostas para a xeración automática (Nazar et al. 2020).

Pezas léxicas	Clases semánticas
santuario   concello   cortello   consulado   abadía	{lugar construción tipo}
cuarto   aposento   vestíbulo   bodega   salón   baño   alcoba	{lugar construción habitación}
cidade   capital   poboación   aldea	{lugar poboación}
Cuba   España   Alemaña   Francia   México	{lugar poboación país nome propio}
Berlín   Roma   Praga   Madrid   París	{lugar poboación cidade nome propio}
monte   outeiro   lomba   montaña   deserto	{lugar paisaxe}
Norte   Sur   Leste   Oeste	{lugar punto cardinal}
provincia   distrito   comarca   finca	{lugar territorio}

Táboa 5. Exemplo de pezas léxicas e clases semánticas na expresión da dirección/destino en gl. fuxida.

Cómpre salientarmos que tanto as clases semánticas como os candidatos léxicos recollidos en cada clase semántica non son categorías universais: a súa categorización está ligada a un substantivo e argumento concreto<sup>26</sup>, que é o que Gross (2008: 11) denomina unha “classe d’objets”. En resumo, o procedemento seguido permítenos describir as preferencias combinatorias dun portador valencial nunha estrutura sintáctico-semántica concreta —o que Bosque denomina “restrición léxica” en REDES—, así como tamén o significado e a súa extensión (Serra Sepúlveda 2009: 205).

Desde un punto de vista metodolóxico e en prol de dotar os xeradores do vocabulario específico máis variado posible asociado a cada clase semántica, nos xeradores séguese ademais un procedemento de expansión léxica a partir de WordNet. O resultado é o que denominamos os paquetes léxicos. Para a súa creación aplicouse en *XeraWord* o mesmo procedemento que na creación dos paquetes para o español, alemán e francés (Domínguez, Solla & Valcárcel 2019), recorréndose, para o galego, ao Galnet<sup>27</sup> (Gómez Guinovart 2011). No caso do portugués seguimos un procedemento de tradución do caudal de WordNet mediante unha ferramenta desenvolvida *ad hoc* para tal fin, *TraduWord* (Domínguez et al., no prelo). Esta ferramenta de tradución automática emprega non só os datos de Wordnet do portugués (PULO<sup>28</sup>; Simões & Gómez Guinovart 2014), senón que estes son combinados con datos procedentes de *MyMemory*<sup>29</sup>. Deste xeito, os paquetes

26. Así, por exemplo, a clase semántica {lugar construción tipo} non recolle o mesmo vocabulario —as mesmas pezas léxicas— no caso de dt. UMZUG (mudanza) que no caso de dt. FLUCHT (fuxida), dado que no primeiro caso son representativas unidades do tipo *casa* | *piso* | *chalé* | *edificio* | *apartamento* etc.

27. <http://sli.uvigo.gal/galnet/> [02/01/2021]

28. <http://www.wordnet.pt/> [15/01/2021]

29. Trátase dunha memoria de tradución colaborativa en liña. Dispoñible en: <https://mymemory.translated.net/> [11/01/2021]

orixinais do español foron traducidos de maneira automática ao portugués e, nun paso posterior, revisados manualmente<sup>30</sup>.

Por tanto, o vocabulario recollido nos xeradores —isto é, os paquetes léxicos— ofrece ao usuario un repertorio de pezas léxicas que son actualizables nunha clase semántica concreta dun argumento determinado dun nome.

## 4. Resultados sobre o emprego dos xeradores para o estudo lingüístico-comparativo

### 4.1. Cuestións introdutorias

Co fin de exemplificar a aplicabilidade dos xeradores para o desenvolvemento de estudos comparativos, así como consultas plurilingües<sup>31</sup>, este apartado presenta algúns aspectos no contraste de gal. FUXIDA cos seus equivalentes en español (*huida*), alemán (*Flucht*) e portugués (*fuga*). Contemplamos esta serie de substantivos<sup>32</sup> como representantes do escenario do desprazamento (§2.1.) por conter no seu padrón argumental un rol focalizado de expresión da dirección, xa for de orixe, paso ou destino (Táboa 4). Por este motivo, imos centrar a nosa atención na realización deste rol semántico e as súas subclases, así como nas clases semánticas asociadas a este.

Tal e como subliñamos no apartado 2.2., esta aproximación, que toma como punto de partida inicial os roles semánticos e non as realizacións, conduce a incluír tamén os adxectivos, os compostos e as construcións N1+N2 (Valcárcel 2017) como realizacións argumentais. Esta cuestión xa nos leva a contemplar unha das diferenzas formais máis salientables entre as tres linguas románicas e o alemán: a actualización de roles semánticos e argumentos en alemán mediante o primeiro elemento léxico de compostos, tal como ilustra a táboa 6.

---

30. Isto por dúas razóns principais: por un lado, para corrixir os posibles erros nas traducións ou eliminar aquelas traducións que non eran as desexadas para un lema concreto (por exemplo, no caso das palabras polisémicas); por outro lado, para comprobar que as unidades léxicas dos paquetes fosen representativas tamén para o portugués, baseándonos nos datos da lingua real extraídos previamente de Sketch Engine.

31. *Xera* e *XeraWord* aínda non están interconectados neste momento e, por tanto, as conclusións ás que pode chegar un usuario das ferramentas resultan da observación independente dos datos.

32. Os datos cuantitativos sobre este paradigma nominal atópanse no apartado 3.1., os esquemas argumentais de FUXIDA no 3.2.1.

	ESTRUTURA	CLASES SEMÁNTICAS
<b>N1</b> Aquel/ Aquilo que realiza a acción	<b>DET + {Adxectivo} + N1 [Animado] + núcleo</b> <b>Die + eilige + Familien- + -flucht</b> [A + apresurada+ familia + fuxida]	{animado humano ideoloxía política} {animado humano axente acto negativo} {animado animal mamífero} {animado humano familia} {animado humano xeral} {animado humano orixe}
<b>N2</b> Locación: orixe	<b>DET + {Adxectivo} + N2 [Locación] + núcleo</b> <b>Die + heimliche + Deutschland- + -flucht</b> [A + secreta+ Alemaña + fuxida]	{lugar construción habitación} {lugar poboación país} {lugar territorio illa} {lugar territorio} {lugar territorio área} {lugar paisaxe} {lugar territorio parte} {lugar construción tipo}
<b>N3</b> Locación: paso	<b>DET + {Adxectivo} + N3 [Locación] + núcleo</b> <b>Die + gefährliche + Mauer- + -flucht</b> [A + perigosa+ muro + fuxida]	{lugar poboación urbanismo canal} {lugar construción parte} {lugar paisaxe acuática} {lugar paisaxe}
<b>N4</b> Locación: destino	<b>DET + {Adxectivo} + N4 [Locación] + núcleo</b> <b>Die + gescheiterte + Schweiz- + -flucht</b> [A + fracasada+ Suíza + fuxida]	{lugar construción habitación} {lugar territorio} {lugar construción tipo} {lugar poboación país} {lugar territorio illa} {lugar paisaxe}

Táboa 6. Compostos argumentais con *-flucht* en alemán.

A escolla da serie de substantivos FUXIDA | HUIDA | FLUCHT | FUGA non é arbitraria. Como é ben sabido, os estudos sobre a expresión do desprazamento no plano oracional son moi numerosos en diferentes linguas (Talmy 2000; Morimoto 2001; Wälchli 2001; Berthele 2006; Slobin 2006; Albi 2010). Se ben dos dicionarios podemos extraer información de diversa índole sobre os substantivos, o número de dicionarios monolingües valenciais (Sommerfeldt & Schreiber 1983; Herbst et al. 2004), así como de gramáticas aplicadas ao ensino de linguas estranxeiras coa información estruturada en campos semánticos e, en xeral, de recursos plurilingües sobre a valencia nominal, aínda é moi reducido. Neste contexto a ferramenta *XeraWord* resulta innovadora, en especial pola facilidade de acceso aos datos. Un usuario pode observar, por exemplo, como se expresa en portugués un {lugar poboación} e que conxunto de candidatos léxicos e realizacións formais adoitan aparecer asociados á devandita clase semántica {lugar poboación}. Deste xeito, a ferramenta ten unha función descritiva, pois permite a xeración dos exemplos

concretos, ordenados por columnas segundo os roles<sup>33</sup> e realizacións (nominais fronte a adxectivos, Figura 6).

**Xera**

Idioma  
Português

Núcleo  
fuga en singular

Filtro semántico  
lugar

Tipo de lugar  
poboación

Límite de frases para xerar: 250

	Anotación semántica	Exemplo
<input type="radio"/> país nome propio a {permanente} fuga {permanente} de Angola	<input type="radio"/> cidade nome propio a {rápida} fuga {rápida} por Alepo	<input type="radio"/> cidade nome propio a {perigosa} fuga {perigosa} até Alexandria
<input type="radio"/> cidade nome propio a {perigosa} fuga {perigosa} de Lisboa	<input type="radio"/> país nome propio a {dramática} fuga {dramática} por Uganda	<input type="radio"/> país nome propio a {dramática} fuga {dramática} até Cuba
<input type="radio"/> cidade nome propio a {constante} fuga {constante} do Rio de Janeiro	<input type="radio"/> país nome propio a {repentina} fuga {repentina} pela Alemanha	<input type="radio"/> cidade nome propio a {vertiginosa} fuga {vertiginosa} até a Havana
<input type="radio"/> poboación a {continúa} fuga {continúa} da capital de Portugal	<input type="radio"/> poboación a {audaciosa} fuga {audaciosa} pela cidade	<input type="radio"/> país nome propio a {repentina} fuga {repentina} até as Bahamas
<input type="radio"/> país nome propio a {importante} fuga {importante} do Brasil	<input type="radio"/> cidade nome propio a {perigosa} fuga {perigosa} pela Cidade do Cabo	<input type="radio"/> poboación a {audaciosa} fuga {audaciosa} até a metrópole
	<input type="radio"/> urbanismo vía a {incrível} fuga {incrível} pela ponte	<input type="radio"/> cidade nome propio a {rápida} fuga {rápida} para Berlín
		<input type="radio"/> país nome propio a {dramática} fuga {dramática} para Cabo Verde
		<input type="radio"/> cidade nome propio a {vertiginosa} fuga {vertiginosa} para o Porto
		<input type="radio"/> país nome propio a {repentina} fuga {repentina} para o Congo
		<input type="radio"/> poboación a {audaciosa} fuga {audaciosa} para a metrópole

Figura 6. Realizacións formais para a clase {lugar poboación} en *XeraWord* para o portugués.

Tendo en conta o anterior, nos seguintes apartados serán tratadas e exemplificadas as cuestións máis relevantes relativas á expresión da dirección de orixe, destino e paso nas linguas obxecto de análise. Para tal fin, o procedemento analítico consiste en seleccionar e agrupar para cada un dos tipos de dirección as súas realizacións de superficie asociadas aos diferentes trazos ontolóxicos e, por conseguinte, ás diferentes clases semánticas. Deste xeito, pódense observar, por unha banda, as restricións ou preferencias combinatorias (por exemplo, a aparición dunha ou outra preposición, a presenza ou ausencia de determinantes) e, por outra banda, as distribucionais (por exemplo, no caso dos adxectivos), tanto dende un punto de vista monolingüe como contrastivo. Son as escenas e as clases semánticas (3.2.2 e 3.2.3) as que permiten a comparación multilingüe.

33. O usuario non visualiza na ferramenta a información sobre os roles semánticos, que si están presentes como elementos centrais na descrición lingüística.

## 4.2. Expresión da dirección de orixe

Da observación dos datos que achegan os xeradores, conclúese que na expresión da dirección de orixe o galego (§3.2.1) e o español (Táboa 7) teñen un comportamento similar, dado que os argumentos do substantivo gl. FUXIDA / esp. HUIDA poden aparecer introducidos polas mesmas preposicións (*de* e *desde*; Táboa 7).

					<b>de</b>	<b>DET</b>	<b>N2</b>	
					de	la	capital	
					<b>de</b>		<b>N2</b>	
					de	Andorra		
<b>N2</b>	<b>Locación:</b>	<b>DET</b>	<b>{Adxectivo}</b>	<b>núcleo</b>	<b>{Adxectivo}</b>	<b>desde</b>	<b>DET</b>	<b>N2</b>
<b>orixe</b>	La	rápida	huida	rápida	desde	el	pueblo	
					<b>de</b>		<b>N2</b>	
					desde	España		

Táboa 7. Estructuras argumentais para Locación: orixe en español.

A diferenza do español, no galego hai variación na preposición, isto é: *desdel dende*. Nesta fase do traballo decidimos seguir un criterio de frecuencia para escoller unha das dúas variantes, criterio que ademais achega información relevante para un usuario que, de seguir a descrición do recursos, estaría a usar a variante máis frecuente no galego actual. A estimación das frecuencias realízase, naturalmente, a partir de datos de corpus: no CORGA podemos ver que a forma máis frecuente é *desde*, con máis de 800 casos por millón, fronte aos preto de 400 de *dende*<sup>34</sup>.

Unha diverxencia interlingüística entre os esquemas argumentais do español e o galego fronte ao portugués tamén se observa no relativo á construción coa preposición *desde*. A pesar de que algunhas gramáticas desta lingua indican a existencia de usos con valor espacial (Raposo et al. 2013), o común é que se empregue con valor temporal. Os datos de corpus para a busca da estrutura *fuga desde* en portugués corroboran a escasa frecuencia de uso desta preposición con valor espacial neste contexto: apenas en 5 dos 32 exemplos que achega Sketch Engine para esta construción a preposición ten valor espacial. O común, así pois, é o uso de *desde* con valor temporal como ilustra o seguinte exemplo<sup>35</sup>:

34. Trátase, non obstante, dunha escolla provisional, xa que o obxectivo futuro é que se poidan integrar na ferramenta este tipo de variacións

35. O exemplo amosa non só a baixa frecuencia da preposición, polo que este tipo de exemplos non están contemplados na descrición da estrutura argumental do substantivo, senón tamén o feito de que neste caso a preposición non está a introducir un actante de FUGA, polo que se trata dun circunstante verbal. É este outro dos motivos para a non inclusión deste tipo de exemplos no recurso.



Dois sospeitos apanhados ficaram em preventiva e dupla está em fuga desde quarta-feira após tentativa de ‘carjacking’. (Sketch Engine)

Outra cuestión relevante está relacionada coas clases semánticas {lugar poboación cidade nome propio} e {lugar poboación país nome propio} e cunha particularidade á hora de establecer os paquetes léxicos<sup>36</sup> para a expresión da dirección nas diferentes linguas: a posible presenza do artigo antes dos nomes de países. No caso do galego, portugués e alemán hai que contemplar que certos nomes de países teñen que ir acompañados de artigo (cando este non forma parte do nome propio, por exemplo, *El Salvador*). A extracción desta información de WordNet para os paquetes non se puido realizar automaticamente, polo que foi preciso compilar dous paquetes diferentes: un para os nomes de países que van precedidos do artigo e outro para os que non.

En canto ao alemán, alén dos xa citados compostos (4.1), dúas preposicións, *von* e *aus*, son centrais na expresión da orixe. A primeira indica o punto de partida do desprazamento e aparece, salvo contadas excepcións, combinada cunha preposición de dirección de destino ou punto final, como, por exemplo, *von Berlin nach Paris* (‘de Berlín a París’). Fronte a *aus*, a preposición *von* tamén pode expresar o axente da acción, por tanto, outro argumento. A segunda, *aus*, indica a procedencia ou orixe no espazo ou no tempo ou a dirección de dentro cara fóra. Cómpre subliñarmos que no uso monoargumental os paquetes combinables con *von* son escasos: é posible a súa realización, por exemplo, nas clases semánticas {lugar territorio illa} e {lugar territorio área}, porén o número de pezas léxicas dos paquetes é moi reducido. A preposición *aus* amosa restricións de combinatoria condicionada á presenza ou ausencia do determinante (Figura 7).

		<i>aus</i> + SUBS + {lugar}	<i>aus</i> + DET + SUBS + {lugar}
poboación	cidade nome propio	+	--
	país nome propio	+	+
territorio	nome propio	+	+
	illa nome propio	+	--
	paisaxe	--	+
	parte	--	+
construción	tipo	--	+
	habitación	--	+
punto cardinal		+	+

Figura 7. Restricións de combinatoria da preposición *aus* en alemán.

36. Un paquete léxico describe un conxunto de unidades léxicas que, non sendo intercambiabes, manteñen entre si una relación paradigmática similar. Conta ademais con información precisa para o seu tratamento e codificación informática (Domínguez & Bardanca & Simões. 2021).

### 4.3. Expresión da dirección de paso

As tres linguas románicas obxecto de análise presentan unha situación similar na expresión da dirección de paso. Podemos tomar como exemplo os esquemas do portugués para identificar os tres esquemas que atopamos nestas tres linguas, xa que a única variación afecta á realización formal da preposición en cada lingua (Táboa 8).

					<b>por</b>	<b>DET</b>	<b>N3</b>
						pelo	mar
<b>N3</b>				<b>{Adxectivo}</b>			
<b>Locación:</b>	<b>DET</b>	<b>{Adxectivo}</b>	<b>núcleo</b>	propositada	<b>por</b>		<b>N3</b>
<b>paso</b>	A	propositada	fuga		por		África
					<b>Adxectivo A3</b>		
<b>A3</b>					marítima		

Táboa 8. Estruturas argumentais para Locación: paso en portugués.

Así, temos por un lado o uso da preposición *por*, seguida ou non de determinante, coas xa citadas particularidades para o caso dos nomes de países e cidades en galego e portugués. En relación cos adxectivos, a realización da clase semántica {lugar paisaxe} en forma adxectival —*a fuxida aérea* (=polo aire), *a fuxida marítima* (=polo mar)— parece específica deste argumento de paso, fronte aos de orixe e destino.

De xeito implícito, tamén se pode observar que existen restricións na distribución. No caso do adxectivo non argumental contéplase unicamente a posición prenominal —*a iminente fuga marítima*, e non *\*a marítima fuga propositada*, *\*a iminente marítima fuga*, *\*a marítima iminente fuga*—, caso contrario ao que sucede coas realizacións preposicionais (onde se contemplan tanto a posición pre- como posnominal).

A diferenza das linguas románicas, o alemán conta, xunto coas realizacións adxectivais e os compostos, con dúas realizacións preposicionais: a frase preposicional con *über* e con *durch*<sup>37</sup>. Un contraste da información recollida nos xeradores leva a observar as diferentes restricións ou preferencias (§3.2.3.) de actualización destas preposicións segundo as clases semánticas (Figura 8).

37. A preposición *über* describe o movemento pasando a través dun lugar cara a un obxectivo máis alá; *durch* focaliza o movemento percorrendo un espazo con entrada e saída, isto é, pasándoo ou atravesándoo dun lado a outro.

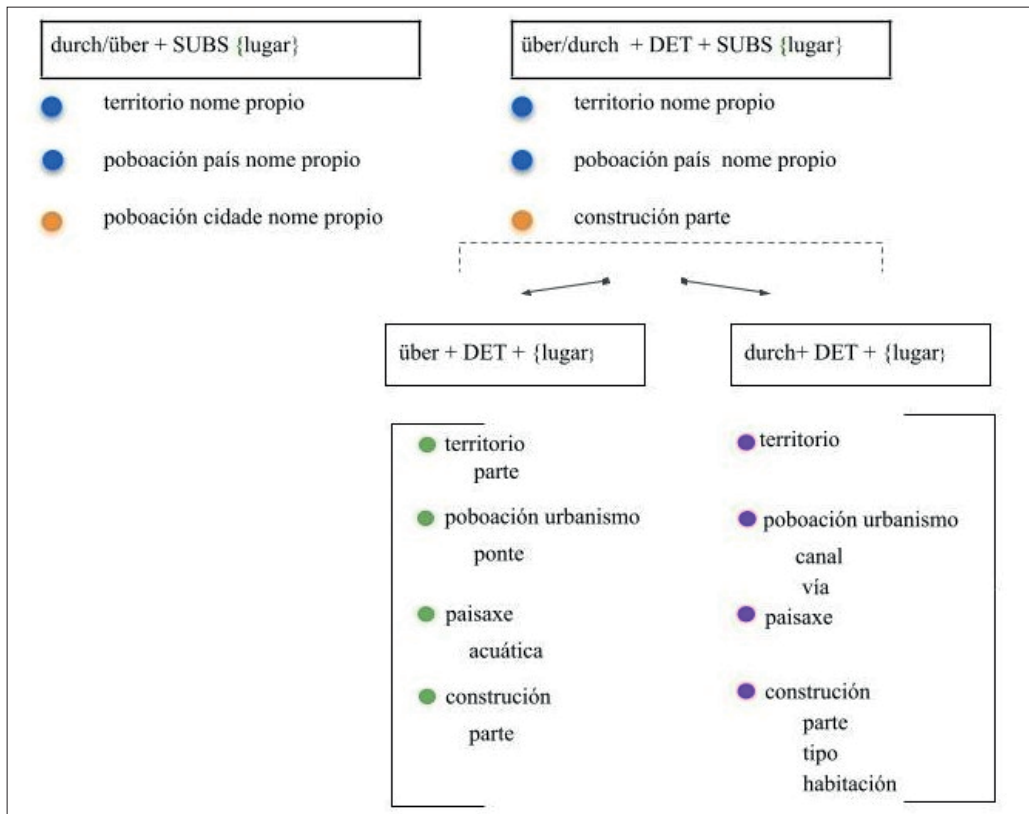


Figura 8. Distribución de clases semánticas na expresión frasal de paso en alemán.

A figura 8 permite resumir unha serie de cuestións relativas ás clases semánticas e á expresión de paso en alemán (Táboa 9).

Determinante <sup>38</sup>	Preposición intercambiable <sup>39</sup>	Marcación
+ e -	+	o
+ ou -	+	o
+	só über	o
+	só durch	o

Táboa 9. Parámetros distintivos para a análise das clases semánticas na expresión frasal de paso en alemán.

Ademais, unha observación detallada da información que achega *XeraWord* para o padrón con determinante fai posible concluír, por unha banda, que existen

38. Fronte a “+ ou -”, “+ e -” indica que ambas opcións –con e sen determinante– son posibles.

39. A marca “intercambiable” non indica que se dilúan as diferenzas semánticas ao usar unha ou outra preposición.

usos restrinxidos de determinados paquetes e preposicións; por outra, que aínda que nalgúns casos pareza haber unha alternancia non restrinxida, este non é o caso: as pezas léxicas non son sempre coincidentes (§3.2.3). Un exemplo deste segundo caso é a clase semántica {construción parte}: en consonancia co seu significado, a preposición *durch* permite introducir aquelas partes dunha construción que podemos atravesar, como unha porta ou unha ventá. No caso de *über* a listaxe de candidatos léxicos redúcese ao vocábulo *muro* e os seus tipos.

#### 4.4. Expresión da dirección de destino

Na expresión da dirección de destino o español sérvese das estruturas introducidas polas preposicións *a*, *hacia* e *hasta*<sup>40</sup>. Se comparamos os esquemas do galego (§3.2.1.) cos do español, podemos establecer unha equivalencia entre os pares *a*[gl]-*a*[es], *ata*-*hasta* e *cara a*-*hacia*. A preposición *a* indica propiamente o destino (*a la capital, a Madrid*), *cara a* e *hacia* fan fincapé na dirección e o par *ata*<sup>41</sup>/*hasta* fan referencia a un límite, que non necesariamente ten que coincidir co punto de destino (Álvarez & Xove 2002: 638).

A particularidade do galego con respecto ao español é a presenza da preposición *para*<sup>42</sup>. A inclusión desta preposición no xerador responde ao seu frecuente uso con valor espacial. Así, Álvarez & Xove (2002: 639) inclúen a preposición *para* entre aquelas cuxo significado básico é a expresión da dirección. Do mesmo xeito, Freixeiro Mato sostén que a preposición *para*, no seu valor principal, “indica dirección ou aproximación espacio-temporal, sendo polo xeral equivalente a *cara a* ou *contra*, mais diferenciándose de *a* por acrecentar certa idea de permanencia no lugar” (Freixeiro Mato 2006: 560-561). Non obstante, este autor tamén menciona outros significados derivados deste significado básico entre os que estarían a expresión de destino ou de aproximación, esta última cando o *para* vai seguido dun adverbio. Por tanto, podería dicirse que esta variación nos significados de *para* fai que, en determinados contextos, esta sexa intercambiable coa preposición *a*, con escasas (ou ningunha) mudanzas no sentido da frase na que se insire.

No portugués contamos con esquemas introducidos polas preposicións *para* (*para o Brasil, para Portugal*) e *até* (*até o Brasil, até Portugal*). A preposición *até* ten un uso similar aos equivalentes nas outras linguas románicas. A diferenza comparativa

40. Algúns exemplos son: *La {repentina} huida {repentina} a la ciudad/la Madrid; La {rápida} huida {rápida} hacia el pueblo/hacia Brasil; La {peligrosa} huida {peligrosa} hasta la calle/hasta Caracas.*

41. Tal e como acontecía no caso da preposición *desde* (§4.2.), *ata* tamén presenta variación en galego (*atalaté*). Nesta ocasión tamén foi seguido un criterio de frecuencia para a escolla da variante *ata*: esta forma conta no CORGA cunha frecuencia de máis de 900 casos por millón, fronte aos máis de 100 casos/millón de *até*.

42. No caso do español a preposición *para* neste contexto estaría a expresar finalidade, e non destino, e non estaría a introducir, polo tanto, un actante do substantivo HUIDA.

máis salientable ten que ver coa preposición *para*, cuxo emprego pode denotar dirección, destino ou límite, cumprindo así as funcións das outras preposicións que se documentan para as outras dúas linguas.

Na lingua alemá tamén atopamos notables restricións na relación entre clases semánticas e selección preposicional, xunto co feito de que, fronte ás linguas románicas, o alemán amosa unha importante variedade de preposicións para a expresión da dirección de destino (Figura 9).

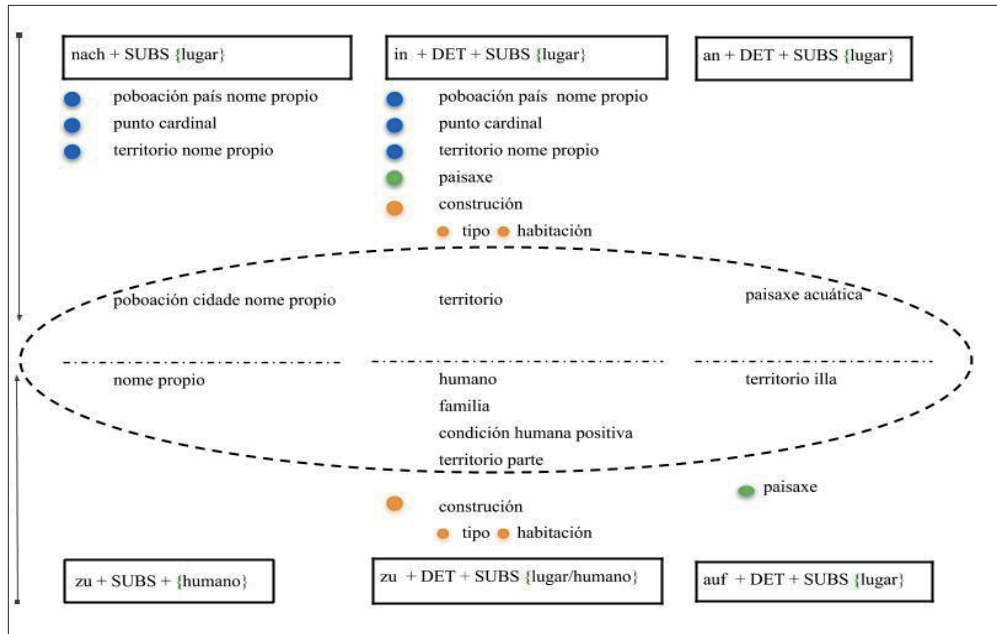


Figura 9: Distribución de paquetes semánticos e preposicións en alemán.

A Figura 9 ilustra as seguintes cuestións:

- A preposición *nach* é a única que non permite a combinación con determinante, xunto con *zu*, no caso de que apareza en combinación cun paquete {humano nome propio}.
- As clases semánticas incluídas no círculo central son aquelas que aparecen exclusivamente cunha determinada preposición. Así, por exemplo, a clase semántica {poboación cidade nome propio} só pode actualizarse coa preposición *nach*, así como todo o relativo aos humanos —ou ao lugar onde estes se atopan— aparece con *zu*.
- Mediante o código de cores diferenciamos diversas casuísticas:
  - A cor azul emprégase para amosar clases semánticas que se actualizan con diferentes preposicións dependendo de se se realiza un determinante ou non.

- A cor laranxa dá conta da intercambiabilidade das preposicións e a conseguinte variación semántica, como é, por exemplo, o caso das preposicións *zu* e *in*. No primeiro caso indícase que o desprazamento se realiza nunha determinada dirección —o que, por exemplo, equivalería a *cara a* en galego ou *hacia* en español—; no segundo caso, con *in*, indícase que de feito se entra nalgún lugar.
- Dado que o vocabulario recollido nos paquetes léxicos dunha mesma categoría semántica non é coincidente (§3.2.3), decidimos a marcación verde unicamente para aqueles casos nos que o paradigma de pezas léxicas é moi limitado. Desta forma, para o caso de {paisaxe} cómpre empregar *in* se nos adentramos nun lugar —como por exemplo a xungla; porén, *auf* + {paisaxe} conta con moi poucos candidatos léxicos, por exemplo, *montaña*, de aí a marcación en verde.

## 5. Conclusións e perspectivas

Os recursos presentados naceron co obxectivo de xerar automaticamente exemplos de padróns argumentais do nome, que permitisen a súa integración noutros recursos en rede ou o seu emprego directo como ferramentas plurilingües. Son, en definitiva, prototipos de dicionarios plurilingües de valencias.

Se ben *Xera* e *XeraWord* comparten a mesma metodoloxía central, o segundo prototipo citado presenta unha estrutura de acceso aos datos especialmente innovadora, dado que —a diferenza de *Xera*— parte de conceptos como as escenas e as clases semánticas e amosa na interface web de consulta *a posteriori* as realizacións de superficie ligadas a estas. Unha vez verificada a viabilidade desta aproximación no novo prototipo *XeraWord* —de aí o seu reducido número de entradas nominais— cumpría unha reflexión sobre as posibles aplicacións e o desenvolvemento futuro dos recursos existentes, en concreto, sobre a posibilidade de empregar os recursos tanto para estudos plurilingües e comparativos como no ensino-aprendizaxe de linguas estranxeiras.

Neste contexto, a consulta monolingüe para o galego e o portugués en *XeraWord*, así como na ferramenta *Xera* para o alemán e español, permite observar as clases semánticas e o caudal léxico que adoita ocupar un determinado espazo funcional —se ben a achega aos datos é no primeiro caso semántico-conceptual e no segundo formal—. Por tanto, ofrecen ao usuario información imprescindible nunha situación de produción ou corrección. Por outro lado, a observación das restricións con respecto ás clases semánticas e as pezas léxicas que as conforman xustifica a metodoloxía aplicada na construción dos xeradores (§3). Ademais, os resultados comparativos presentados ao longo do apartado 4 xustifican a necesidade de acometer estudos intralingüísticos,



mais tamén inter e multilingüísticos, certamente escasos para o contraste entre o alemán e o galego, por exemplo. Estes estudos, á súa vez, poden ter unha aplicación directa no ensino de linguas estraxeiras. As ferramentas de xeración automática achegan información valencial cuxa aprendizaxe pode carrexar certas dificultades (§2.1.), especialmente cando non se dispón dos recursos necesarios, como é o caso. Por tanto, unha ferramenta plurilingüe *cross lingual* —como o dicionario Portlex (2018)— que ofrezca dun xeito amigable para o usuario este tipo de información tería unha notable utilidade no ámbito do ensino de linguas.

A modo de resumo final, cómpre remarcarmos que a optimización dos recursos pasa pola retroalimentación entre os xeradores e a integración dos seus datos, entre estes e outras ferramentas, así como pola automatización nos procedementos descritivos. Comprobada a viabilidade metodolóxica de *XeraWord*, este terá que ser ampliado e perfeccionado mediante a descrición de máis lemas e a introdución da combinatoria biargumental, que xa está en funcionamento para outras linguas, como exemplifica a ferramenta *Combinatoria*. Finalmente é preciso subliñar que estes recursos de xeración automática están sendo integrados nunha ferramenta máis ampla, con acceso conceptual-semántico aos datos —seguindo o modelo aquí presentado de *XeraWord*— que implementará ademais o marco oracional e que, por conseguinte, potenciará a posibilidade de ser empregada como recurso didáctico online para o galego, español, francés, alemán e portugués.

## Agradecementos

Quixeramos agradecer ao equipo de revisión anónimo as súas valiosas achegas e comentarios.

## Financiamento

Esta investigación enmárcase nos proxectos “Ferramentas TraduWord e XeraWord: tradución de caudal léxico e xeración automática da linguaxe natural en galego e portugués” (*XeraWord*; programa “Convocatoria de proxectos colaborativos para institutos de investigación da USC. 2020”), “Generación multilingüe de estruturas argumentales del sustantivo y automatización de extracción de datos sintáctico-semánticos” (MultiGenera; Ayudas Fundación BBVA a Equipos de Investigación Científica 2017) e “Generador multilingüe de estructuras argumentales del sustantivo con aplicación en la producción en lenguas extranjeras” (MultiComb; MINECO; Programa Excelencia, FFI2017-8245,4-P).

## Referencias bibliográficas

- Albi Aparicio, Miguel Á. 2010. *La valencia lógico-semántica de los verbos de movimiento y posición en alemán y español*. Frankfurt a.M.: Lang.
- Álvarez Blanco, Rosario & Xosé Xove. 2002. *Gramática da lingua galega*. Vigo: Galaxia.
- Berthele, Raphael. 2006. *Ort und Weg. Die sprachliche Raumreferenz in Varietäten des Deutschen, Rätoromanischen und Französischen*. Berlin & New York: de Gruyter. <https://doi.org/10.1515/9783110890464>
- Busse, Winfried. 1994. *Diccionario sintáctico de verbos portugueses*. Coimbra: Almedina.
- Busse, Dietrich. 2012. *Frame Semantik. Einführung, Diskussion, Weiterentwicklung*. Berlin & Boston: De Gruyter.
- Cidrás Escáneo, Francisco Antonio. 2017. Determinación nominal con posesivo en español y gallego. En María José Domínguez Vázquez & Silvia Kutscher (eds.), *Interacción entre gramática, didáctica y lexicografía*. 239-250. Berlin & Boston: De Gruyter. <https://doi.org/10.1515/9783110420784-018>
- Combinatoria* = Domínguez Vázquez, María José (dir.), Carlos Valcárcel Riveiro, Daniel Bardanca Outeiriño, Natalia Catalá Torres, Nerea López Iglesias, Rosa Martín Gascueña, Mónica Mirazo Balsa, María Teresa Sanmarco Bande & Laura Pino Serrano. 2020. *Combinatoria. Prototipo online para la generación biargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela. <http://portlex.usc.gal/combinatoria/usuario> [11/01/2021].
- CORGA = Centro Ramón Piñeiro para a investigación en humanidades: *Corpus de Referencia do Galego Actual (CORGA) [3.2]* <http://corpus.cirp.gal/corga/> [09/01/2021].
- DAELE = *Diccionario de Aprendizaje de Español como Lengua Extranjera*: <http://www.iula.upf.edu/recl/dae/> [15/01/2021].
- DEPE = Maldonado, Concepción. (dir.). (2002). *Diccionario de español para extranjeros*. Madrid: SM.
- DICE = Alonso Ramos, Margarita. 2004. *Diccionario de colocaciones del español (DICE)*. <http://www.dicesp.com/paginas> [02/01/2021].
- Domínguez Noya, Eva María, M.<sup>a</sup> Sol López Martínez & Francisco Mario Barcala Rodríguez. 2019. Corpus de Referencia do Galego Actual (CORGA): composición, codificación, etiquetaxe e explotación. En Marta Blanco, Hella Olbertz & Victoria Vázquez Rozas (coords.), *Corpus y construcciones: perspectivas hispánicas*. 179-218. Santiago de Compostela: Universidade de Santiago de Compostela. <https://dx.doi.org/10.15304/9788417595876>
- Domínguez Vázquez, María José. 2015. Wie wird Veränderung im Deutschen und im Spanischen ausgedrückt? Ein szenen- und valenzfundiertes konzeptuelles wortklassenübergreifendes Beschreibungsmodell. En Meike Meliss & Bernhard Pöll (eds.), *Aktuelle Perspektiven der kontrastiven Sprachwissenschaft: Deutsch-Spanisch-Portugiesisch. Zwischen Tradition und Innovation*. 97-125. Tübingen: G. Narr.
- Domínguez Vázquez, María José. 2020. Aplicación de WordNet e de word embeddings no desenvolvemento de prototipos para a xeración automática da lingua. *Linguamática* 2, 71-80. <https://doi.org/10.21814/lm.12.2.337>

- Domínguez Vázquez, María José, Daniel Bardanca & Alberto Simões. 2021. Automatic Lexicographic Content Creation for Lexicographers. En: Elex 2021: *Post-editing Lexicography. Proceedings of the eLex 2021 conference*. 5-7 July 2021. Brno: Lexical Computing CZ s.r.o. [No prelo].
- Domínguez Vázquez, María José & Carlos Valcárcel Riveiro. 2020. PORTLEX as a multilingual and cross-lingual online dictionary. En María José Domínguez Vázquez, Mónica Mirazo Balsa & Carlos Valcárcel Riveiro (eds.), *Studies on multilingual lexicography*. 135-158. Berlin: De Gruyter. <https://doi.org/10.1515/9783110607659-008>
- Domínguez Vázquez, María José, Miguel Anxo Solla Portela & Carlos Valcárcel Riveiro. 2019. Resources interoperability: exploiting lexicographic data to automatically generate dictionary examples. En Iztok Kosem & Tanara Zingano Kuhn (eds.), *Proceedings of the VI. eLex conference Electronic lexicography in the 21st century: Smart Lexicography*. 51-71. Brno: Lexical Computing CZ s.r.o.
- Domínguez Vázquez, María José, Alberto Simões, Daniel Bardanca Outeiriño, María Caíña Hurtado & José Luis Iglesias Anllones. Automatic Generation of Nominal Phrases for Portuguese and Galician Combining Multilingual Resources into *XeraWord*. [No prelo]
- DRAG = Real Academia Galega. 2021. *Dicionario da Real Academia Galega*. <https://academia.gal/dicionario>. [01/04/2021].
- Engel, Ulrich. 1996. Semantische Relatoren. Ein Entwurf für künftige Valenzwörterbücher. En Nico Weber (ed.), *Semantik, Lexikographie und Computeranwendung*. 223-236. Tübingen: Niemeyer. <https://doi.org/10.1515/9783111555522.223>
- Engel, Ulrich. 2004. *Deutsche Grammatik – Neubearbeitung*. München: Iudicium.
- Freixeiro Mato, Xosé Ramón. 2006. *Gramática da lingua galega*. Vigo: A Nosa Terra.
- Gao, Jun & Haitao Liu. 2020. Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners. *Lexikos* 30, 111-142. <https://doi.org/10.5788/30-1-1548>
- Fillmore, Charles John. 1977. Scenes-and-Frames Semantics. En Antonio Zampolli (ed.), *Linguistic Structures Processing*. 55-81. Amsterdam, New York & Oxford: North-Holland.
- Fuertes-Olivera, Pedro, Marta Niño Amo & Ángeles Sastre Ruano. 2019. Tecnología con fines lexicográficos. *RILE. Revista Internacional de Lenguas Extranjeras*, 10, 75-100. <https://doi.org/10.17345/rile10.75-100>
- Gómez Guinovart, Xavier. 2011. Galnet: WordNet 3.0 do galego. *Linguamática* 3/1, 61-67. <https://doi.org/10.17979/rgf.2015.16.0.1383>
- Gross, Gaston. 2008. *Les classes d'objets*. Paris: Presses de l'Ecole normale supérieure.
- Hanks, Patrick (2004). Corpus Pattern Analysis. En Geoffrey Williams & Sandra Vessier (eds.), *Proceedings of the XI EURALEX International Congress*. 87-97. Lorient: Université de Bretagne-Sud.
- Hanks, Patrick & James Pustejovsky. (2005). A Pattern Dictionary for Natural Language Processing. *Revue Française de Linguistique Appliquée* 10(2), 63-82. <https://doi.org/10.3917/rfla.102.82>
- Herbst, Thomas, David Heath, Ian. F. Roe & Dieter Götz. 2004. *A Valency Dictionary of English: a corpus based analysis of the complementation patterns of English verbs, nouns and adjectives*. Berlin: de Gruyter. <https://doi.org/10.1515/9783110892581>
- Herger, Hans Jürgen. 1984. Neues von der Verbszene. En Gerhard Stickel (ed.), *Pragmatik in der Grammatik: Jahrbuch 1983 des Instituts für deutsche Sprache*. 34-64. Düsseldorf: Schwann.
- Hermanns, Fritz & Werner Holly. 2007. *Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens*. Berlin: de Gruyter.

- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography ASIALEX* 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kipper, Karin, Hoa Trang Dang & Martha Palmer. 2000. Class-based construction of a verb lexicon. En Robert Englemore & Haym Hirsh (eds.), *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*. 691–696. Austin: AAAI Press.
- Laufer, Batia & Paul Nation. 2012. Vocabulary. En Susan M. Gass & Alison Mackey (eds.), *The Routledge Handbook of Second Language Acquisition*. 163–176. London/New York: Routledge.
- Meyers, Adam, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, & Ralph Grishman. 2004. Annotating noun argument structure for NomBank. En Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, Raquel Silva, Carla Pereira, Filipa Carvalho, Milene Lopes, Mónica Catarino & Sérgio Barro (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*. París: European Language Resources Association (ELRA).
- Morimoto, Yuko (2001): *Los verbos de movimiento*. Madrid: Visor.
- Müller-Spitzer, Carolin, María José Domínguez Vázquez, Martina Nied Curcio, Idalete Maria Silva Dias & Sascha Wolfer. 2018. Correct Hypotheses and Careful Reading Are Essential: Results of an Observational Study on Learners Using Online Language Resources. *Lexikos* 28, 287-315. <https://doi.org/10.5788/28-1-1466>
- Müller-Spitzer, Carolin, Martina Nied Curcio, María José Domínguez Vázquez, Idalete Maria Silva Dias & Sascha Wolfer. 2019. Recherchepraxis bei der Verbesserung von Interferenzfehlern aus dem Italienischen, Portugiesischen und Spanischen: Eine explorative Beobachtungsstudie mit DaF-Lernenden. *Lexicographica* 34, 157-182. <https://doi.org/10.1515/lex-2018-0008>
- Nazar, Rogelio, Antonio Balvet, Gabriela Ferraro, Rafael Marin & Irene Renau. 2020. Pruning and repopulating a lexical taxonomy: experiments in Spanish, English and French. *Journal of Intelligent Systems* 30 (1), 376-394. <https://doi.org/10.1515/jisys-2020-0044>
- Nied Curcio, Martina. 2014. Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht. En Andrea Abel, Chiara Vettori & Natascia Ralli (eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 263-280. Bolzano & Bozen: Institute for Specialised Communication and Multilingualism.
- Pérez Serrano, Mercedes. 2014. ¿Son indispensables los diccionarios combinatorios? *Revista de lexicografía* 20, 121-145. <https://doi.org/10.17979/rlex.2014.20.0.3877>
- Portlex = Domínguez Vázquez (dir.), Carlos Valcárcel Riveiro, Mónica Mirazo Balsa, María Teresa Sanmarco Bande, Alberto Simões & Mário João Vale. 2018. *Portlex. Diccionario multilingüe de la valencia del nombre*. Universidade de Santiago de Compostela. [http://portlex.usc.gal/portlex/\[01/01/2021\]](http://portlex.usc.gal/portlex/[01/01/2021]).
- Raposo, Eduardo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura & Amália Mendes (coord.) 2013. *Gramática do português*. Lisboa: Fundação Calouste Gulbenkian.
- REDES = Bosque, Ignacio (dir.) 2004. *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: SM. <https://doi.org/10.17979/rlex.2007.13.0.4792>

- Ruppenhofer, Josef, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R. Johnson, & Jan Scheffczyk. 2006. *FrameNet II: Extended theory and practice*. Berkeley: International Computer Science Institute.
- Sánchez Rufat, Anna & Francisco Jiménez Calderón. 2013. Combinatoria léxica y corpus como input. *Language Design* 14, 61-81.
- Serra Sepúlveda, Susana. 2009. Las restricciones de selección en los diccionarios generales de lengua española. *Boletín de Filología* 44 (2), 187-213. <https://doi.org/10.4067/s0718-93032009000100007>
- Schumacher, Helmut. 2006. Deutschsprachige Valenzwörterbücher. En Vilmos Ágel, Ludwig M. Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependenz und Valenz. Ein internationales Handbuch der zeitgenössischen Forschung, vol. II*. 1396–1424. Berlin & New York: De Gruyter. <https://doi.org/10.1515/9783110171525.2.11.1396>
- Simões, Alberto & Xavier Gómez Guinovart. 2014. Bootstrapping a portuguese wordnet from galician, spanish and english wordnets. En Juan Luis Navarro Mesa, Alfonso Ortega, António Teixeira, Eduardo Hernández Pérez, Pedro Quintana Morales, Antonio Ravelo García, Iván Guerra Moreno & Doroteo T. Toledano (eds.), *Advances in Speech and Language Technologies for Iberian Languages*. 239-248. Berlin: Springer International Publishing. <https://doi.org/10.1007/978-3-319-13623-3>
- Slobin, Dan I. 2006. What makes manner of motion salient? En Maya Hickmann, & Stéphane Robert, S. (eds.): *Space in languages: Linguistic systems and cognitive categories*. 59-82. Amsterdam & Philadelphia: Benjamins. <https://doi.org/10.1075/tsl.66.05slo>
- Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.
- Sommerfeldt, Karl-Ernst & Herbert Schreiber. 1983. *Wörterbuch zur Valenz und Distribution der Substantive*. Leipzig & Tübingen: De Gruyter. <https://doi.org/10.1515/9783111549491>
- Talmy, Leonard. 2000. *Towards a Cognitive Semantics, Volume II: Typologie and Process in Concept Structuring*. Cambridge: MIT Press.
- VALBU = Schumacher, Helmut, Jacqueline Kubczak, Renate Schmidt & Vera de Ruiter. 2004. *VALBU - Valenzwörterbuch deutscher Verben*, Tübingen: Narr. <http://hypermedia.ids-mannheim.de/evalbu/index.html> [02/02/2021].
- Valcárcel Riveiro, Carlos. 2017. Las construcciones N1N2 como realizaciones actanciales del sustantivo en francés y su tratamiento en el diccionario multilingüe PORTLEX. En María José Domínguez Vázquez & Silvia Kutscher (eds.), *Interacción entre gramática, didáctica y lexicografía*. 193-207. Berlin & Boston: De Gruyter. <https://doi.org/10.1515/9783110420784-015>
- Wälchli, Bernhard. 2001. A typology of displacement (with special reference of Latvian). *Sprachtypologie und Universalforschung* 53/3, 298-323. <https://doi.org/10.1524/stuf.2001.54.3.298>
- VLdr = Engel, Ulrich & Emilia Savin,. 1983. *Valenzlexikon deutsch-rumänisch. Dictionar de valenta german-român*. Heidelberg: Groos.
- VLdsp = Domínguez Vázquez, María José, Ulrich Engel & Gemma Paredes Suárez. 2017. *Neue Wege zur Verbvalenz. Band II: Deutsch-spanisches Valenzlexikon*. Frankfurt: Peter Lang.
- Xera = Domínguez Vázquez, María José (dir.), Carlos Valcárcel Riveiro, Daniel Bardanca Outeiriño, José Antonio Calañas, Natalia Catalá Torres, Rosa Martín Gascuña, Mónica Mirazo Balsa, María Teresa Sanmarco Bande & Laura Pino Serrano. 2020. *Xera. Prototipo online para la generación automática monoargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela. <http://portlex.usc.gal/combinatoria/usuario> [11/01/2021].

*XeraWord* = Domínguez Vázquez, María José (dir.), Daniel Bardanca Outeiriño, María Caíña Hurtado, Xavier Gómez Guinovart, José Luis Iglesias Allones, Alberto Simões, Carlos Valcárcel Riveiro, María Álvarez de la Granja & Francisco A. Cidrás Escaneo. 2020. *XeraWord. Prototipo de xeración automática da argumentación da frase nominal en galego e portugués*. Santiago de Compostela: Instituto da Lingua Galega. <http://ilg.usc.gal/xeraword/> [02/01/2021]

Ziem, Alexander. 2008. *Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz*. Berlin & New York: De Gruyter. <https://doi.org/10.1515/9783110209419>