

TECHLING 2021

UVIGO — T&P



Universidade de Vigo



OS PROBLEMAS SEMÁNTICOS NO CORPUS DE TEXTOS GALEGOS ESCRITOS POR ESTUDANTES NO ÁMBITO ACADÉMICO (CORTEGAL)

María Álvarez de la Granja
Instituto da Lingua Galega-Universidade de Santiago de Compostela

Esta presentación é parte do proxecto de I+D+i
PGC2018-096069-B-I00, financiado/a por MCIN/
AEI/10.13039/501100011033/ e FEDER
“Una manera de hacer Europa”



MINISTERIO
DE CIENCIA
E INNOVACIÓN

UNIÓN EUROPEA



FONDO EUROPEO DE
DESARROLLO REGIONAL
“Una manera de hacer Europa”



Obxectivo

- Analizar os problemas semánticos no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL) e intentar establecer unha clasificación explicativa dos usos semánticos non estándares encontrados



Textos de CORTEGAL

- Corpus conformado por 1000 textos redactados en galego por estudantes de segundo de Bacharelato (17-18 anos) da comunidade autónoma galega
- Redaccións elaboradas como resposta á pregunta número 3 dos exames da materia Lingua e literatura galegas da proba de Avaliación do Bacharelato para o acceso á Universidade (ABAU) (curso 2016-2017)
- Na pregunta 3, solicítase que o/a estudante redacte un texto de *ao redor de 200-250* palabras en que reflicta *de forma argumentada a súa opinión* sobre un tema determinado
- As formas non estándares presentes nos textos de CORTEGAL están sendo anotadas (con asignación da forma estándar correspondente e cun código identificador do tipo de problema) en seis niveis lingüísticos (ortográfico, morfolóxico, léxico, gramatical, semántico e discursivo)



Problemas semánticos en CORTEGAL

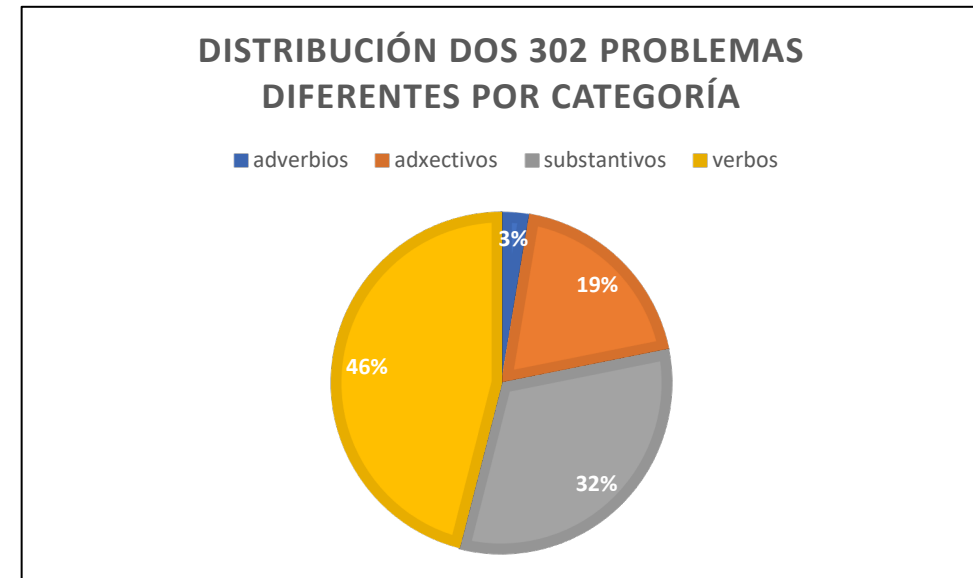
- 400 primeiros textos anotados en CORTEGAL
- Palabras ou expresións cuxos significados ou usos en CORTEGAL non se corresponden cos significados ou empregos convencionais ou establecidos normativamente
- Substantivos, adxectivos, verbos e adverbios



Resultados

- 451 problemas semánticos (1,13 por texto)
 - O problema máis frecuente (35 ocorrencias) é o emprego do verbo *facer* con valor temporal, correspondente a *haber* no galego estándar:
- 302 problemas diferentes e no 90% dos casos con só unha ocorrencia

“prezos impensables *fai* un par de anos”



Orixe dos problemas

- Análise de carácter explicativo: causas do emprego dunha forma semanticamente inapropiada no contexto en que aparece ou do uso dunha forma non estándar
- Subxectividade
 - “apoio na bioloxía para defender que outro modelo produtivo non tería cabida na nosa *raza*”



Orixe dos problemas

- **Calcos semánticos non aceptados no estándar:** adquisición por parte dunha unidade léxica galega dun novo significado por influencia dunha palabra doutra lingua coa que a primeira (a palabra galega) mantén vínculos formais e/ou semánticos
- Do castelán: *fac*er (por *haber*), *pob*o (por *vila, aldea...*), *desenrol*ar (por *desenvolver...*), *contempl*ar (por *considerar*) etc.
- Do inglés: *pretend*er (por *finxir*), *balanc*e (por *equilibrio*)
- Moitos deles son calcos habituais en galego, que son etiquetados en CORTEGAL por razóns meramente prescritivas.



Orixe dos problemas

- **Outros significados orixinados no propio galego pero non aceptados no estándar**
- *verba*: 'palabra': "unha imaxe vale por mil *verbas*"
(DRAG: 'Habilidade para agradar e convencer con palabras')



Orixe dos problemas

- **Ampliacións semánticas de palabras por influencia de sinónimos parciais**
- *derradeiro*: “Nos *derradeiros* anos a gastronomía a dado un gran paso adiante a nivel cultural”
- *réxime*: “Á hora de realizar unha viaxe non só é importante a diversidade dos monumentos senón que tamén importa a comida. En parte, a comida e o tipo de alimentación tamen forman parte da cultura. Pódese aprender moito analizando os tipos de *réximes* levados a cabo nos diversos países”



Orixe dos problemas

- **Ampliacións semánticas de palabras mediante a aplicación de mecanismos produtivos**
- *popularidade*: “programas televisivos con diferentes **popularidades** gastronómicas ao seu frente” (*metonimia*)
- *salientar*: “a gastronomía comezou a **salientar** entre os temas de interese público” (*intransitivización*)



Orixe dos problemas

- **Lapsus na selección explicables por varios motivos**
- *Presenza no contexto lingüístico próximo da palabra empregada erroneamente, ou dalgunha palabra do seu mesmo campo léxico:*

“En segundo lugar en canto aos prezos dos utensilios que mercamos cada día deben achegarse ao **prezos** que dispoñen as familias para poder aumentar as vendas en todos os sectores e non perder o volumen das vendas senón gañalo”



Orixe dos problemas

- **Lapsus na selección explicables por varios motivos**
- *Selección errónea dalgunha das moitas palabras que se activan no proceso de elaboración do texto, vinculadas co campo temático que este aborda*

“Cada vez no noso modelo de vida, consumimos máis produtos e **compramos** menos facendo que cada vez sexa máis difícil manter un bo ritmo de produtividade que aporte beneficios ós fabricantes dos produtos”



Orixes dos problemas

- **Lapsus na selección explicables por varios motivos**
- *Confusión e mestura entre dúas posibles opcións para expresar aproximadamente o mesmo contido*

“Mais vale a pena pagar o prezo que *teñen* por esa comida”



Orixe dos problemas

- **Lapsus na selección explicables por varios motivos**
- *Confusión cunha palabra próxima formalmente*

“O consumo está presente na nosa vida dende o noso nacemento, cando nos envolven nunha manta que foi comprada con anterioridade, e dende alí, somos consumidores sempre. Dentro disto hai xente máis *costumista* que outra, pero na nosa sociedade actual todo ten un custo, e temos que comprar o necesario é a base do mundo capitalista”





Orixe dos problemas

- **Descoñecemento por parte da /do estudante do significado ou uso exacto da palabra ou expresión**

“Imos ao supermercado e compramos cousas innecesarias porque nos chama a atención. Pero todas esas cousas non nos fan falta realmente. Coa metade das cousas poderíamos vivir perfectamente. Sería máis *factible* non gastar tanto diñeiro en produtos que non necesitamos e axudarlle a xente que realmente o necesita pero que esta non ten medios para mercalo”



Orixe dos problemas

- **Descoñecemento por parte da /do estudante do significado ou uso exacto da palabra ou expresión**
- *Coñecemento aproximado do seu valor*

“É moi significativa a importancia do consumismo e a produtividade na nosa vida. A variación nunha ter *represarias* na outra, unha leva a outra”.

“unha gran cantidade do prezo que pagas e simplemente que leve a *insignia* da empresa”



Orixe dos problemas

- **Descoñecemento por parte da /do estudante do significado ou uso exacto da palabra ou expresión**
- *Confusión con outra palabra próxima formalmente*

“Nestes días, a crise económica é *latente* nos ámbitos sociais e económicos”



Algúns casos frecuentes

- **Selección de substantivos xenéricos abstractos**

“Dende fai uns anos a cociña e a gastronomía vai tendo unha maior popularidade, onde pouco a pouco valórase máis estes *feitos*”

“os cidadáns navegaban cara Occidente en busca de materiais que intercambiaban por outra cousa que interesara aos vendedores, chamado así este *fenómeno* economía de troque”

“Outro *ámbito* que cabe resaltar da gastronomía é a súa presenza nos medios de comunicación”

“delegamos como conxunto de cidadáns a nosa soberanía en quen permite que sucedan estos *asuntos* que son denunciados no texto: EREs, baixos salarios...”



Algúns casos frecuentes

- **Selección de palabras, nomeadamente colocados, que serven para expresar valores xenéricos do tipo de 'existencia', 'presenza', 'aparición', 'creación', 'posta en práctica'...**
 - “A cociña nos últimos anos *sufriu* un gran crecemento”
 - “este sistema capitalista no que estamos inmersos *presenta* unha gran influencia na sociedade”
 - “ao mesmo tempo *orixinamos* traballo aos produtores deses utensilios”
 - “*formación* de purés”
 - “Aquí moitas persoas *obteñen* curiosidade ou aprecio pola gastronomía de outros lugares”



Algúns casos frecuentes

- **Selección de adxectivos e adverbios con función ponderativa**

“crise do sistema inmobiliario que vivimos *fortemente* en España hai uns anos”

“non sabemos o que pode ocorrer no futuro e por iso non debemos afirmar nada *completamente*”

“A gastronomía e a cociña acadaron unha *forte* popularidade estos últimos anos”

“non se olvidan da importancia do sabor, polo que tamén mostran un *importante* interese”

“Sí que é certo que hay un *gran* factor social e tratase probablemente dunha moda”



Interese didáctico de CORTEGAL

Xerador de consultas

Busca de texto

palabra	comeza por ▼	feito
Estandarización ortográfica	igual a ▼	
Estandarización morfolóxica	igual a ▼	
Estandarización léxica	igual a ▼	
Estandarización gramatical	igual a ▼	
Estandarización semántica	igual a ▼	
Estandarización discursiva	igual a ▼	
Tipo de problema	S_lu_su ▼	

Mostrar contexto que non sexa simplemente | un **feito** para satisfacer unha necesidade vital

Mostrar contexto , non só polo **feito** | das inmensas listas de

Mostrar contexto xa | que este é un **feito** con moita presenza na

Mostrar contexto moda duradeira polo meiro **feito** da | cantidade

Mostrar contexto antes a vían como o **feito** de alimentarse para | sobrevi

Mostrar contexto por exemplo Master Chef. | O **feito** deste interese pola

Mostrar contexto satisfacer ou polo **feito** de necesidade. | | Vivimos n

Mostrar contexto pouco valórase [...] **máis** estes **feitos** , considerándoos como aficio

Mostrar contexto como a cociña son dous **feitos** moi importantes na nosa

Mostrar contexto e máis a produción son **feitos** que están presentes | constantemente na

Mostrar contexto igual, | xa que istos **feitos** teñen cousas positivas pero tamén

Mostrar contexto e a consumición son dous **feitos** que creceron | exponencialmente. Pero

feito	
Estandarización semántica	actividade
Lema estándar	feito
Clase de palabra (estándar)	Nome (NCMS000) Common; masculine; singular
Tipo de problema	S_lu_su





Interese didáctico de CORTEGAL

A gastronomía e a cociña acadaron unha forte popularidade estes últimos anos debido a tendencia das persoas a probar cousas novas en canto a comida e o aumento da competición neste sector. Cada día os cociñeiros experimentan con novos tipos de pratos, especies, presentacións, etc, que lle da certo atractivo as persoas que lle gusta probar novos pratos. Ademais, para moitos é considerado un novo arte xa que non só o sabor é o importante, tamén o é a presentación do prato, que debe ser delicada e coidadosa, a mezcla dos elementos combinados, as texturas.... Isto provocou unha competición por ver qué prato era o máis novidoso; o máis sabroso, o que intensificou todavía máis o aumento da gastronomía no mundo, xa que a competición era de carácter mundial. Pronto comezaron os títulos de gastronomía que nomeaban ao mellor gastrónomo según diferentes graos de superioridade e os programas de televisión. Estes últimos foron un claro avance no mundo gastronómico, xa que o dotaba de máis importancia. Actualmente, hai moitos programas dedicados a elaboración de comidas, e de competicións por ver cal é o mellor, donde tamén participan nenos pequenos, xa que hai programas dedicados especialmente para eles.

Esta nova situación do mundo da cociña, que se deixou de ver como algo específico para mulleres, supuxo un cambio cultural na forma de elaborar as comidas e na interpretación dos pratos, chegando a convertirse nun novo arte e nunha profesión prestixiosa.

A gastronomía e a cociña acadaron unha gran popularidade estes últimos anos debido á tendencia das persoas a probar pratos novos e ao aumento da competición neste sector. Cada día os cociñeiros experimentan con novos tipos de pratos, especies, presentacións etc., que lle dan certo atractivo ao sector entre as persoas a que lles gusta probar novos pratos. Ademais, para moitos a gastronomía é considerada unha nova arte, xa que non só o sabor é o importante, tamén o é a presentación do prato, que debe ser delicada e coidadosa, a mestura dos elementos combinados, as texturas.... Isto provocou unha competición por ver que prato era o máis novo, o máis sabroso, o que intensificou aínda máis o crecemento da gastronomía no mundo, xa que a competición era de carácter mundial. Pronto comezaron os títulos de gastronomía, que nomeaban os mellores gastrónomos segundo diferentes graos de mestría, e os programas de televisión. Estes últimos supuxeron un claro avance no mundo gastronómico, xa que o dotaban de máis importancia. Actualmente, hai moitos programas dedicados á elaboración de comidas e de competicións por ver cal é o mellor, onde tamén participan nenos pequenos, xa que hai programas dedicados especialmente a eles.

Esta nova situación do mundo da cociña, que se deixou de ver como algo específico para mulleres, supuxo un cambio cultural na forma de elaborar as comidas e na interpretación dos pratos, chegando a converterse nunha nova arte e nunha profesión prestixiosa.



Bibliografía



techLING'21

- Abel, Andrea et al. (2014). "KoKo: an L1 Learner Corpus for German", en N. Calzolari et al. (eds.), [*Proceedings of the Ninth International Conference on Language Resources and Evaluation \(LREC-2014\)*](#). Reykjavik: European Languages Resources Association, 2414-2421. http://www.lrec-conf.org/proceedings/lrec2014/pdf/934_Paper.pdf
- Abel, Andrea / Aivars Glaznieks / Lionel Nicolas / Egon Stemle (2016): "An extended version of the KoKo German L1 Learner corpus", en Anna Corazza / Simonetta Montemagni / Giovanni Semeraro (dirs.), *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016*. Torino: Accademia University Press. <http://ceur-ws.org/Vol-1749/paper1.pdf>
- Abel, Andrea Aivars Glaznieks (2017): *KoKo: Bildungssprache im Vergleich: korpusunterstützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum - ein Ergebnisbericht*. Bolzano: Eurac Resarch 1.0 <<https://hdl.handle.net/10863/7500>>
- Amaro, Raquel / Correia, Susana / Gramacho, Carolina / Mendes, Amália (2020): "Automatização no diagnóstico de nível de língua: anotação e versatilidade dos recursos para PLE", *Revista da Associação Portuguesa de Linguística* 7, 1-20 <https://doi.org/10.26334/2183-9077/rapln7ano2020a1>



Bibliografía

- Anctil, Dominic (2005): *Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et analyse des stratégies de résolution de problèmes lexicaux*. Montréal: Université de Montréal.
- Boyd, Adriane / Hana, Jirka / Nicolas, Lionel / Meurers, Detmar / Wisniewski, Katrin / Abel, Andrea / Schone, Karin / Stindlov, Barbora / Vettori, Chiara (2014): "The MERLIN corpus: Learner language and the CEFR", *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 14)*, 26-31 May 2014, pp.1281-1288;
- Dagneaux, Estelle / Sharon Denness / Sylviane Granger (1998): "Computer-aided error analysis", *System* 26, 126-174.
- [Díaz-Negrillo](#), Ana / [Jesús Fernández Domínguez](#) (2006): "Error Tagging Systems for Learner Corpora", [Revista española de lingüística aplicada](#) 19, 83-102.
- González Álvarez, Elsa (1999): "Análisis de los errores léxico-semánticos", en Luis Iglesias Rábade (coord.), *Análisis de los errores del examen de inglés en las pruebas de acceso a la Universidad en el distrito universitario de Galicia*. Santiago de Compostela: Instituto de Ciencias da Educación - Universidade de Santiago de Compostela, 207-270.

Bibliografía



tecLING'21

Instituto Galego de Estatística. (2019): *Enquisa estrutural a fogares. Coñecemento e uso do galego*.

Santiago de Compostela: Instituto Galego de Estatística

<https://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0206004>.

Janssen, Maarten (2016): "[TEITOK: Text-Faithful Annotated Corpora](#)", en Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asunción Moreno, Jan Odijk, & Stelios Piperidis (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož: European Language Resources Association (ELRA), 4037-4043. http://www.lrec-conf.org/proceedings/lrec2016/pdf/651_Paper.pdf

Lüdeling, Anke / Hagen Hirschmann (2015): "Error annotation systems", en Sylviane Granger / Gaëtanelle Gilquin / Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press, 135-158.

Lüdeling, Anke, Walter, Maik, Kroymann, Emil, Adolphs, Peter (2005): "Multi-level error annotation in learner corpora", *Proceedings from the Corpus Linguistics Conference Series*, 1 (1). Birmingham: University of Birmingham <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>



Bibliografía

- Mendes, Amália, Sandra Antunes, Maarten Janssen & Anabela Gonçalves (2016): “The COPLE2 Corpus: A Learner Corpus for Portuguese”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož: European Language Resources Association (ELRA), 3207-3214.
- MERLIN project (2014): *Annotation guidelines*. <https://merlin-platform.eu/docs/Annotation%20guidelines.pdf>.
- Reznicek, Marc Lüdeling, Anke, Hirschmann Hagen (2013): “Competing target hypotheses in the Falko corpus. A flexible multi-layer corpus architecture”, en Ana Díaz-Negrillo, Nicolas Ballier, Paul Thompson (eds.), [*Automatic Treatment and Analysis of Learner Corpus Data*](#) [[Studies in Corpus Linguistics](#) 59], pp. 101-124.
- Reznicek, M., Ludeling, A., Krummes, C., Schwantuschke, F., Walter, M., Schmidt, K., Hirschmann, H. & Andreas, T. (2012): *Das Falko-Handbuch: Korpusaufbau und Annotationen. Version 2.01*. Berlin: Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/view>



Bibliografía

- Río, I. del / A. Mendes (2018): “Error annotation in the COPLE2 corpus”, *Revista da Associação Portuguesa de Linguística* 4, 225-239 <https://doi.org/10.26334/2183-9077/rapln4ano2018a42>.
- Stemle, Egon / Boyd Adriane / Janssen Maarten / Mikelić Preradović, Nives / Rosen, Alexandr / Rosén, Dan / Volodina, Elena (2019). “Working together towards an ideal infrastructure for language learner corpora”, en Andrea Abel, Aivars Glaznieks, Verena Lyding, Lionel Nicolas (eds.), *Widening the Scope of Learner Corpus Research. Selected papers from the fourth Learner Corpus Research Conference*. Louvain-la-Neuve: Presses universitaires de Louvain, 427-468.
- Tenfjord, Kari / Meurer, Paul / Hofland, Knut (2006): “The ASK corpus: A language learner corpus of Norwegian as a second language”, en *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, 24-26, 1821–1824.
<https://pdfs.semanticscholar.org/af74/17fdebb1d51e08f4b77111d6464fbaa9f6a5.pdf?_ga=2.246502660.1873985775.1585827674-334967235.1585669811>
- Wisniewski, Katrin / Woldt, Claudia / Schöne, Karin / Abel ,Andrea / Blaschitz, Verena / Štindlová, Barbara / Vodičková, Kateřina (2014): *The MERLIN annotation scheme for the annotation of German, Italian, and Czech learner language*. www.merlin-platform.eu

MOITAS GRAZAS!!!

María Álvarez de la Granja
Instituto da Lingua Galega-Universidade de Santiago de Compostela
maría.alvarez.delagranja@usc.gal

Esta presentación é parte do proxecto de I+D+i
PGC2018-096069-B-I00, financiado/a por MCIN/
AEI/10.13039/501100011033/ e FEDER
“Una manera de hacer Europa”



MINISTERIO
DE CIENCIA
E INNOVACIÓN

