

***Perspectivas sobre a
escrita académica en
galego***

***Estudos baseados no corpus
CORTEGAL de textos
escritos por estudantes***

**Edición a cargo de
María Álvarez de la Granja
Vítor Míguez**

Verba

Anexo 84

2023

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PERSPECTIVAS SOBRE A ESCRITA
ACADÉMICA EN GALEGO
ESTUDOS BASEADOS NO CORPUS *CORTEGAL* DE
TEXTOS ESCRITOS POR ESTUDANTES

Edición a cargo de
MARÍA ÁLVAREZ DE LA GRANJA
VÍTOR MÍGUEZ

Verba
Anexo 84

2023

Universidade de Santiago de Compostela



Esta obra atópase baixo unha licenza internacional Creative Commons BY-NC-ND 4.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-ND 4.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.gl>

A publicación deste libro foi posible grazas á colaboración do Ministerio de Cultura e Deporte do Goberno de España mediante subvención ao Instituto da Lingua Galega. A elaboración do corpus *CORTEGAL* en que se basean os traballos foi posible grazas ao apoio financeiro da Secretaría Xeral de Política Lingüística da Xunta de Galicia a través de convenios de colaboración co Instituto da Lingua Galega e grazas ao financiamento do proxecto PGC2018-096069-B-I00 por parte do Ministerio de Ciencia e Innovación e Universidades do Goberno de España.

© Universidade de Santiago de Compostela, 2023

Maquetación

Tania Sanmartín

Edita

Edicións USC

Campus Vida

15782 Santiago de compostela

usc.gal/publicacions

DOI <https://dx.doi.org/10.15304/9788419679970>

ISSN 2341-1198

ÍNDICE

- 05 **Introdución**
MARÍA ÁLVAREZ DE LA GRANJA, VÍTOR MÍGUEZ
- 19 ***CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico. Deseño do corpus e caracterización dos textos***
MARÍA ÁLVAREZ DE LA GRANJA, REYES RODRÍGUEZ RODRÍGUEZ
- 93 **Análise da acentuación gráfica a través do corpus *CORTEGAL***
MARÍA LÓPEZ-SÁNDEZ, LARA LORENZO-HERRERA
- 115 **O léxico non estándar en *CORTEGAL***
MARÍA ÁLVAREZ DE LA GRANJA
- 203 **A posición do pronome persoal átono con formas verbais finitas no corpus *CORTEGAL***
XOSÉ ANTONIO FERNÁNDEZ SALGADO
- 253 **Sempre concordamos co que pensamos. Unha aproximación á análise da concordancia baseada nos erros**
FRANCISCO CIDRÁS
- 287 **Procedementos de xeneralización e marcado da implicación no corpus *CORTEGAL*: entre a linguaxe académica e o rexistro coloquial da lingua**
CARME SILVA-DOMÍNGUEZ
- 323 **A expresión da incerteza nos textos de *CORTEGAL*: un achegamento paradigmático**
VÍTOR MÍGUEZ
- 351 **Índice de Táboas**
- 354 **Índice de Gráficas**
- 355 **Índice de Figuras**

Introdución

MARÍA ÁLVAREZ DE LA GRANJA

Instituto da Lingua Galega-Universidade de Santiago de Compostela

VÍTOR MÍGUEZ

Universidad del País Vasco / Euskal Herriko Unibertsitatea

O volume que presentamos, *Perspectivas sobre a escrita académica en galego. Estudos baseados no corpus CORTEGAL de textos escritos por estudantes*, recolle un conxunto de traballos que analizan os textos de *CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico* (Álvarez de la Granja e González Seoane, s.d.). A confección de *CORTEGAL* e do volume partiu de diferentes traballos que, dun xeito ou outro, serviron de fonte de inspiración e de coñecemento. Por tal motivo, abrimos este limiar cunha breve presentación destes antecedentes e de *CORTEGAL*, a cal serve para enmarcar a do propio volume, cuxos capítulos serán introducidos con máis detalle no apartado §2.

1. *CORTEGAL* E OS TRABALLOS PREVIOS

CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico (<http://ilg.usc.gal/cortegal/>), é un corpus de acceso libre elaborado no Instituto da Lingua Galega da Universidade de Santiago de Compostela que contén 1000 producións extraídas do exame de Lingua Galega e Literatura das Probas de avaliación de bacharelato para o acceso á universidade (ABAU), concretamente da pregunta de comentario, en que o alumnao ten que elaborar un texto de carácter argumentativo de entre 200 e 250 palabras sobre un tema determinado, vinculado cun texto previo.

O obxectivo deste volume é ofrecer un conxunto de estudos que, no marco metodolóxico da lingüística de corpus e a partir da análise dos textos de *CORTEGAL*, nos presenta unha análise da escritura académica en lingua

galega do alumnado de Galicia unha vez que remata a educación secundaria. Tales estudos, precedidos dun capítulo que amosa tanto a metodoloxía empregada na elaboración do corpus como as principais características dos textos que o conforman, mostran diferentes posibilidades de achegamento ás producións escritas do alumnado galego.

Trátase do primeiro volume sobre a escrita académica en galego elaborado no marco da lingüística de corpus. Debemos indicar, con todo, que contamos con traballos previos que analizan a destreza da expresión escrita en lingua galega do estudiantado de Galicia, moi particularmente desde a perspectiva da súa avaliación. As publicacións máis relevantes a este respecto derivaron do encargo que lle fixo a Secretaría Xeral de Política Lingüística da Xunta de Galicia ao Instituto de Ciencias de Educación da Universidade de Santiago de Compostela para deseñar e implementar probas que permitisen obter información sobre a competencia instrumental do alumnado galego ao rematar a Educación Secundaria Obrigatoria, tanto en galego como en castelán e nas catro destrezas (Silva Valdivia, 2010, p. 7). Os principais resultados desta investigación, levada a cabo en 2009, recóllense na obra publicada en 2010 *Avaliación da competencia do alumnado de 4º da ESO nos idiomas galego e castelán*, dirixida por Silva Valdivia. Neste proxecto traballouse con probas escritas sobre o coñecemento gramatical e léxico nas dúas linguas, así como cunha proba de produción oral en cada lingua e con outra de produción escrita —a redacción dun texto—. En Loredo Gutiérrez e Silva Valdivia (2014), onde se traballa cos resultados obtidos en Silva Valdivia (2010), analízanse con detalle cales son as principais variables que inflúen no nivel atinxido nas competencias léxica e gramatical. Máis recentemente, en Loredo e Silva Valdivia (2020), obra editada pola Real Academia Galega e que recolle os resultados das probas realizadas no curso 2018-2019, partindo dos mesmos presupostos teóricos e metodolóxicos básicos, revisouse e actualizouse o deseño aplicado en 2009. Nos dous proxectos indicados, o foco sitúase na comparativa co español e na análise da influencia de diferentes variables sociodemográficas, educativas e sociolingüísticas nos resultados encontrados.

Con anterioridade a estes traballos e no marco dun proxecto cun obxectivo similar, «Análise das competencias bilingües ao final do ensino obrigatorio na Comunidade Autónoma de Galicia», dirixido por Manuel González González, desenvolveuse unha proba piloto para avaliar a competencia en galego e castelán do alumnado de 4º da ESO da comarca compostelá. En Loredo Gutiérrez et al. (2005) explícase a metodoloxía desta proba en relación

coa produción escrita, consistente na elaboración dunha redacción sobre un tema previamente seleccionado, e analízase a influencia de distintas variables nos resultados obtidos. Noia (1980), pola súa banda, utiliza diferentes probas (entre elas redaccións) para avaliar a competencia oral e escrita de alumnado de dúas parroquias de Vigo. Outros traballos abordan competencias específicas e utilizan probas de diferente tipo, como no caso de Esteban Radío (1997), que analiza exclusivamente o ámbito léxico a través de fichas de vocabulario, ou no de Silva Valdivia (2006), que estuda os castellanismos morfosintácticos no galego a través de probas de corrección, completión e tradución.

Os traballos mencionados mostran a existencia de certos problemas na produción escrita en lingua galega do estudantado de Galicia e aqueles publicados antes da posta en marcha de *CORTEGAL* serviron en boa medida de estímulo e fonte de ideas para a elaboración deste corpus e para o seu deseño no marco metodolóxico da «Análise informatizada de erros» ou «Análise de erros asistida por ordenador» (Dagneaux et al., 1998), que ten como un dos seus obxectivos centrais a estandarización e codificación das formas ou secuencias que se desvían do estándar para facilitar a súa posterior análise. Deste xeito, tomando como referencia e punto de partida outros corpus elaborados neste marco, que son, sobre todo, corpus de aprendentes de L2 e LE (por exemplo, o corpus *MERLIN*, 2014; o *Falko*, Reznicek et al., 2012; ou o *COPLE2*, Mendes et al., 2016), pero tamén de L1 (como o corpus *Koko*, Abel e Glaznieks, 2017), deseñamos e elaboramos *CORTEGAL*, levando a cabo a anotación das formas non estándar en seis niveis lingüísticos (ortográfico, morfolóxico, léxico, semántico, sintáctico e discursivo) e publicando o primeiro corpus galego desenvolvido no marco metodolóxico que acabamos de mencionar. Así pois, aínda que *CORTEGAL* permite coñecer de maneira xeral as particularidades da escritura académica en galego, non é menos certo que, na liña dos traballos mencionados, un dos obxectivos fundamentais á hora da súa elaboración foi o de poder avaliar a competencia do alumnado galego na destreza da expresión escrita e nun rexistro formal académico, o que explica a anotación das formas desviantes do estándar.

Por outra banda, a pesar das dificultades e limitacións que supón o emprego de textos extraídos das probas ABAU, expostas no primeiro capítulo desta obra, optamos por traballar con esta fonte pola significancia e representatividade de tales probas para a sociedade galega, tal e como se expón con máis detalle no citado capítulo do volume. Outros traballos, para outras linguas, tamén se achegaron a textos extraídos deste tipo de fonte co obxectivo

de coñecer a competencia na lingua escrita do estudiantado, como é o caso de Iglesias Rábade (1999), onde, no marco da análise de erros (Corder, 1967; Spillner, 2017), se analiza, en diferentes niveis lingüísticos, un corpus de traballo constituído por exames de inglés das probas de acceso á universidade de Galicia. De igual xeito, en Battaner Arias e Torner Castells (2005) recóllese un conxunto de estudos sobre os textos do corpus PAAU 1992, así como os propios textos, producións escritas en español extraídas de exames de diferentes materias das probas de acceso á universidade de seis comunidades autónomas¹. Díez-Bedmar (2011), de súa vez, conformou tamén un corpus de textos das probas de inglés de Jaén, levando a cabo análise informatizada de erros e nesta mesma obra poden consultarse outros traballos que estudaron textos en inglés extraídos das probas de acceso á universidade.

O interese na avaliación da competencia condúcenos inexorablemente á dimensión didáctica de *CORTEGAL*, que se manifesta por dúas vías distintas. Por unha banda, consideramos que o recurso pode ser empregado directamente nas aulas para mellorar a competencia na destreza da escritura, e que isto pode facerse de diferentes maneiras (explotando as capas de visualización do corpus, as cales permiten ver os textos na súa forma orixinal e nas versións estandarizadas en diferentes niveis lingüísticos; explotando os datos das concordancias a través da aprendizaxe guiada por datos; ou seleccionando textos de interese para traballar co alumnado, por exemplo a partir dos datos cuantitativos que ofrece o buscador). Por outra banda, a análise dos textos de *CORTEGAL* e das súas anotacións por parte das e dos lingüistas permite detectar, describir e explicar os problemas existentes na destreza da escritura no rexistro académico, paso previo necesario para que posteriormente se deseñen estratexias didácticas que conduzan a unha mellora na competencia do alumnado.

Neste sentido, e aínda que varios dos traballos que conforman este volume se centran nas formas desviantes do estándar, debe terse en conta que a obra non ten como obxectivo principal ofrecer propostas didácticas, senón describir e explicar os problemas encontrados. Non obstante, e dado o innegable interese que esta análise ten desde o punto de vista do ensino do galego, os diversos capítulos que conforman a obra incorporan, en maior ou menor medida, reflexións didácticas sobre os problemas atopados, reflexións que, canda a análise sobre os fenómenos estudados, poden supoñer un punto de partida para o deseño das correspondentes medidas didácticas.

¹ Os textos están dispoñibles en <http://www.iula.upf.edu/rec/corpus92/>

2. OS CAPÍTULOS

O volume que presentamos ofrece unha pequena mostra de *perspectivas sobre a escrita académica en galego*, conxugando capítulos que analizan as formas que se desvían do estándar no nivel ortográfico, léxico e gramatical, con outros traballos que nos axudan a coñecer o emprego que o alumnado realiza de determinados mecanismos discursivos. A este respecto, debemos sinalar que, se un dos nosos obxectivos é avaliar a competencia do alumnado na destreza da escritura e no rexistro académico, debemos estudar a tal fin aspectos específicos da competencia discursiva (que, innegablemente, son de gran relevancia), pero tamén aspectos vinculados coa competencia lingüística (Canale, 1980), incluíndo a competencia ortográfica, a gramatical e a léxico-semántica, posto que os discursos se constrúen con unidades léxicas e con relacións gramaticais e porque a adecuación ao rexistro académico pasa, en moi boa medida, polo emprego de ortografía, léxico e gramática estándar. Esta idea presidiu a selección de traballos que conforman a obra, dos que a seguir ofrecemos unha breve presentación.

María Álvarez de la Granja e Reyes Rodríguez Rodríguez abren o volume cun capítulo centrado no deseño de *CORTEGAL* e na caracterización dos seus textos. As autoras comezan describindo a mostra de producións escritas que constitúen o corpus. Como se adiantou ao comezo desta introdución, *CORTEGAL* está composto por 1000 textos manuscritos de entre 200 e 250 palabras correspondentes ao exercicio de comentario do exame de Lingua Galega e Literatura da proba de acceso á universidade. Nas primeiras páxinas do capítulo atopamos unha descrición detallada deste exercicio (incluíndo unha indicación dos temas e opcións correspondentes a cada convocatoria), a distribución dos textos en relación cos temas, as convocatorias e as Comisións Delegadas, así como información sobre a cualificación das probas. A seguir, Álvarez de la Granja e Rodríguez Rodríguez presentan o proceso de transcripción, *tokenización* e anotación dos textos na plataforma TEITOK, prestando especial atención aos criterios que guiaron a transcripción e á asignación dos metadatos. As autoras detéñense nun dos aspectos máis interesantes de *CORTEGAL*: o tratamento das formas non estándar. *CORTEGAL* é un corpus de aprendentes que incorpora análise informatizada de erros e, polo tanto, identifica as formas desviantes do estándar mediante un conxunto de etiquetas e asígnalles as formas estándar correspondentes. Esta tarefa, complexa de seu, foi levada a cabo tendo en conta seis niveis taxonómicos, de maneira que unha mesma forma desviante pode recibir formas normalizadas diferentes en varios niveis, como é o caso de *esten*, a que lle corresponde *estén* no

nivel ortográfico e *estean* no nivel morfolóxico. Ademais da análise de erros, *CORTEGAL* ofrece unha clasificación dos conectores textuais presentes no corpus, como primeiro paso na anotación de aspectos discursivos deste tipo. Este capítulo non só presenta a metodoloxía en que se baseou a elaboración do corpus, senón que tamén achega os seus principais resultados cuantitativos, atendendo, por unha banda, á caracterización xeral dos textos e, por outra banda, aos problemas detectados na análise de erros. Así, obtemos información sobre aspectos como o número total de palabras, a densidade léxica dos textos, a lonxitude dos enunciados, a media de formas desviantes por texto ou a distribución porcentual dos diferentes tipos de erros.

No seguinte capítulo, María López-Sández e Lara Lorenzo-Herrera analizan a acentuación gráfica dos textos de *CORTEGAL*. A acentuación gráfica concentra un alto número de anotacións no corpus, que as autoras entenden como mostra da persistencia dos problemas de acentuación entre o alumnado galego. López-Sández e Lorenzo-Herrera analizan en primeiro lugar os problemas de acentuación segundo se trate de casos de omisión, adición ou desprazamento. Obsérvase aquí un gran desequilibrio entre as tres categorías, sendo a omisión a claramente dominante e o desprazamento un tipo de problema acentual moi raro. As autoras combinan criterios cuantitativos e cualitativos na súa análise e, deste xeito, chegan á conclusión de que a orixe maioritaria dos problemas de omisión e adición é a interferencia entre o sistema de acentuación do galego e o do castelán. No caso da omisión do acento, apoia esta idea o número elevado de casos de acento omitido en palabras con acentuación diacrítica e en formas verbais con clítico, moito máis infrecuentes en castelán. Na adición tamén son frecuentes os problemas coa acentuación diacrítica e destacan outros casos como a acentuación de palabras agudas con ditongo decrecente, dos adverbios en *-mente* ou de certas formas monosílabas. Desde a perspectiva da estrutura acentual da palabra, o maior número de problemas dáse nas palabras agudas, o cal deriva da lóxica do sistema de acentuación gráfica, que, por motivos de economía, acentúa unha maior porcentaxe de palabras agudas que de palabras graves, por seren estas as máis frecuentes na lingua. López-Sández e Lorenzo-Herrera tamén abordan a dispersión dos erros entre o alumnado e observan que a maioría de textos do corpus contén algún tipo de problema de acentuación. Aínda que poucos textos conteñen un número avultado de erros, as autoras sinalan que a ampla presenza destes en exercicios correspondentes ás probas de acceso á universidade é un motivo para a revisión das metodoloxías didácticas.

María Álvarez de la Granja, no terceiro capítulo do volume, achégase ás numerosas formas léxicas non estándar presentes nos textos de *CORTEGAL*

combinando a perspectiva cuantitativa coa cualitativa. Esta contribución comeza presentando o sistema de clasificación de formas léxicas non estándar do corpus, que contén tres tipos básicos: a atribución a un substantivo dun xénero non estándar (por exemplo, **a leite*), a colocación do acento de intensidade nunha sílaba diferente da acentuada na forma estándar (por exemplo, **Nóbel*) e a substitución dunha expresión estándar por unha unidade non estándar con diferenzas que van máis alá do xénero ou da acentuación (por exemplo, **plato*). Ademais, a anotación do corpus adoita indicar a orixe da forma non estándar (por exemplo, coincidencia co español, hipergaleguismo ou analoxía con outras voces galegas). Desde o punto de vista cuantitativo, Álvarez de la Granja mostra que o máis común nos textos do corpus é conter unha forma léxica non estándar, aínda que os textos con entre 2 e 4 ou ningunha son tamén moi frecuentes. As 3025 expresións non estándar corresponden a 787 formas distintas, con diferenzas notables no número de ocorrencias, pois unha forma como **plato* aparece 103 veces no corpus, mentres que hai 434 formas que só ocorren unha vez. Segundo o tipo de problema, a clase máis numerosa é a das diferenzas que van máis alá do xénero ou da acentuación, que concentra o 95% das formas léxicas non estándar, e, segundo a orixe do problema, as formas con influencia directa ou indirecta do español acumulan o 90% dos casos. A análise cualitativa pon o foco en exemplos concretos correspondentes ás tres clases xa mencionadas e ofrece unha panorámica das características dos problemas do ámbito léxico atopados no corpus. De maneira semellante a López-Sández e Lorenzo-Herrera, Álvarez de la Granja apunta que as explicacións máis probables para moitas das formas léxicas non estándar de *CORTEGAL*, dado o contexto en que ten lugar a produción dos textos, son o descoñecemento, a inseguridade ou os lapsos e sinala a necesidade de reforzo do traballo cos aspectos léxicos nas aulas.

Xosé Antonio Fernández Salgado examina a seguir a colocación do pronome persoal átono con formas verbais finitas, un trazo que caracteriza o galego con respecto ás variedades románicas próximas. O traballo ofrece unha visión global deste aspecto da sintaxe pronominal contrastando os datos de *CORTEGAL* coas descrições que ofrecen as gramáticas para a variedade estándar. O 90% dos casos analizados son coincidentes coa posición canónica do galego estándar, mentres que as ocorrencias restantes corresponden, por unha banda, a énclices e próclises non canónicas ou agramaticais e, por outra banda, a outros casos diverxentes do estándar, isto é, colocacións non recomendadas na variedade estándar pero gramaticais noutras variedades da lingua. O capítulo realiza unha análise detallada de casos de énclice canónica, próclise canónica, énclices e próclises non canónicas e outras colocacións

diverxentes do estándar, atendendo sempre a diferentes contextos sintácticos e a factores relevantes como a presenza de elementos que favorecen o movemento do clítico á posición preverbal. Na opinión do autor, o número de casos de colocación do clítico con algún tipo de diverxencia co estándar é reducido. No conxunto de casos anómalos, son máis frecuentes as colocacións enclíticas por proclíticas (por exemplo, **este factor é determinante, xa que ninguén plantéxa<se> o feo que pode ser algo*) que as colocacións proclíticas por enclíticas (por exemplo, **Estas dúas accións <as> leva facendo o ser humano case dende a súa aparición*). Casos do primeiro tipo poden considerarse ultracorreccións, pois a posición posverbal é a non marcada en galego, e poden derivar dunha mala aprendizaxe do galego como segunda lingua, mentres que os casos do segundo tipo son produto da interferencia do castelán. Non obstante, hai que ter en conta que existen factores que fan máis ou menos probable a anomalía, sendo os contextos negativos onde menos se dan as colocacións non canónicas e os contextos con cuantificadores focalizados os que máis as favorecen.

No seguinte capítulo, Francisco Cidrás aborda os erros de concordancia en *CORTEGAL* como ferramenta positiva para levar a cabo unha análise sintáctica que reflexione sobre como funciona o sistema lingüístico. O autor defende que entender a motivación das escollas de concordancia problemáticas favorece a construción do propio coñecemento gramatical, que deste xeito pode ser aplicado á mellora da produción e comprensión lingüísticas. A concordancia mostra unha considerable variabilidade e no seu uso conflúen factores de natureza diversa (morfosintáctica, semántica e pragmática), o cal fai dela unha área gramatical de gran relevancia para a reflexión sintáctica. Ademais, é un dos principais mecanismos para a marcaxe de relacións gramaticais, xunto á marca de caso e a orde de palabras. Porén, a función primixenia da concordancia non é a marcaxe de relacións gramaticais no ámbito da oración, senón a de recurso cohesivo no nivel textual. Esta subsidiariedade da función gramatical a respecto da textual explica que a marca de concordancia se sitúe no verbo, que resulta a opción natural para un mecanismo cohesivo, mais non así para un recurso de marca de relación gramatical. Isto, sumado ás características do funcionamento da propia concordancia, fai que sexa o mecanismo de marcaxe de relacións gramaticais con maior risco potencial de erro. No plano empírico, estas consideracións son apoiadas polos datos que ofrece *CORTEGAL*, onde case a metade dos problemas de natureza sintáctica corresponden a problemas de marcado das relacións gramaticais e, dentro destes, os erros de concordancia acumulan a porcentaxe máis elevada.

Con todo, é necesario diferenciar a concordancia verbal da nominal, pois só no caso da primeira son pertinentes as consideracións anteriores, xa que a concordancia nominal si responde de maneira específica a unha estratexia de marcase das relacións gramaticais, aínda que limitada ao ámbito da frase. Cidrás sinala que a inmensa maioría dos erros de concordancia son produto dun desaxuste entre a conceptualización semántica e a codificación gramatical, como a que se dá nun substantivo como *xente*, con morfoloxía singular e semántica de pluralidade. Os problemas de selección inadecuada de número son os máis frecuentes, con moita diferenza, aínda que tamén son abundantes os relativos ao xénero e, en menor medida, á persoa. Tras estudar con detenimento cada unha destas tres coordenadas da concordancia, o autor remata a súa contribución cunha reflexión sobre as posibilidades de explotación didáctica do tema.

No penúltimo capítulo, Carme Silva-Domínguez estuda os procedementos de xeneralización e marcado da implicación en *CORTEGAL*, é dicir, o xeito de expresar «a relación entre lugar común e punto de vista do/da autor/a». Isto inclúe a expresión tanto de perspectivas compartidas entre un número variado de persoas como de pareceres totalmente individuais, alén dun repertorio de posibilidades intermedias. Dada a natureza dos textos que compoñen *CORTEGAL*, o esperable é que predominen os recursos que fundamentan o punto de vista da persoa emisora en opinións compartidas, de maneira que se vexan favorecidos os mecanismos de xeneralización fronte aos de implicación. Porén, a realidade textual é máis complexa. A implicación dáse nos verbos mediante o uso da primeira persoa do singular, así como da primeira do plural e da segunda do singular en postulados xerais. A terceira persoa do plural e as estruturas pasivas son recursos de xeneralización presentes no corpus, aos que hai que sumar elementos nominais ou pronominais como *a xente*, *todos*, *a sociedade*, *as persoas*, *a poboación*, *o ser humano*, *todo o mundo* ou *ninguén*, que revelan unha preferencia por elementos propios do rexistro coloquial. Estes recursos mostran un comportamento diverxente no tocante á concordancia de persoa. Así, por exemplo, *a xente* ten un uso minoritario coa primeira persoa de plural, mentres que *todos* combina con esta persoa maioritariamente. Este tipo de feitos mostran a complexidade do posicionamento en textos argumentativos, onde atopamos recursos que combinan a xeneralización coa implicación de maneiras sofisticadas. No relativo á linguaxe non sexista, aínda que formas problemáticas como *o home*, *os cidadáns* ou *todos* non son moi usadas no corpus, chama a atención a case total ausencia de formas ideadas para evitar a masculinización do discurso, como *homes e*

mulleres ou *todos e todas*. Silva-Domínguez remata a súa contribución achegando unhas pautas para o traballo escolar dos recursos estudados, non só desde unha perspectiva gramatical, senón tendo en conta tamén o concepto de adecuación.

Vitor Míguez pecha o volume cun traballo sobre os recursos empregados para expresar incerteza en *CORTEGAL*, onde se inclúen verbos auxiliares como *poder*, predicados adxectivos como *é posible*, verbos cognitivos como *crer* e adverbios como *quizais*. Estes constitúen formas alternativas para a expresión dun mesmo contido e, polo tanto, son susceptibles de seren estudados desde unha perspectiva paradigmática. Así, Míguez pretende explicar a variación formal nesta coordenada gramatical estudando os factores que motivan a escolla entre variantes alternativas para a expresión da incerteza. Neste sentido, o autor considera nove variables, entre as cales figuran a (inter)subxectividade, a performatividade, a estratexia discursiva e as relacionadas coa estrutura informativa. Os resultados do estudo mostran que os catro tipos de expresións examinados corresponden a tres perfís funcionais diferentes, constituíndo adverbios e verbos auxiliares variantes dun mesmo perfil. Os predicados adxectivos representan o tipo de expresión menos frecuente e máis especializado, pois posúen unha alta especificidade semántica e están vinculados á modificación adicional da cualificación epistémica (isto é, toman modificadores na maioría das súas ocorrencias), o cal se asocia á centralidade informativa do xuízo de incerteza, que adoita ser información secundaria. Os verbos cognitivos son expresións de dúbida subxectiva e, segundo o autor, a súa prevalencia explícase polo carácter subxectivo do xénero textual argumentativo, ademais de pola súa relevancia como temas modais. Os adverbios e os auxiliares son os recursos que menor sensibilidade mostran aos factores da estrutura informativa e os máis flexibles desde un punto de vista funcional, pois permiten tanto lecturas subxectivas como intersubxectivas, ademais de poder desempeñar función temática ou integrarse no enunciado. A partir do estudo das formas que expresan incerteza, o autor chega a dúas conclusións sobre os textos argumentativos no ámbito académico: por unha banda, as persoas autoras outorgan un papel central á expresión das propias hipóteses e opinións na construción do texto e, por outra banda, a ocorrencia frecuente de expresións de incerteza obedece, na maioría de casos, á necesidade de formular hipóteses e non á consecución de obxectivos estratéxicos como atenuar a forza das propias afirmacións.

Para pechar esta epígrafe, cómpre sinalar que a maior parte dos exemplos numerados que se ofrecen nos diferentes capítulos que conforman o

volume foron extraídos de *CORTEGAL*. As poucas excepcións existentes están convenientemente sinaladas, ben sexa nas liñas previas, ben sexa mediante a indicación da fonte no propio exemplo. Por outra banda, e dado que os exemplos proporcionados se corresponden ás veces con enunciados completos e outras con fragmentos, para lograr a maior homoxeneidade posible na súa presentación, optamos por eliminar o punto final naqueles que non están truncados.

3. CABO

Os capítulos que conforman esta obra mostran como a lingüística de corpus pode facer unha contribución significativa ao estudo da escrita académica en galego e amosan ao mesmo tempo que esta achega pode ser especialmente importante no ámbito da didáctica, posto que pon de relevo os problemas que se encontran nas producións escritas en lingua galega do estudantado de Galicia, sexa no nivel ortográfico, no léxico, no gramatical, no semántico ou no discursivo². Neste sentido, a análise levada a cabo nos diferentes traballos que conforman a obra, estivesen ou non focalizados nas formas desviantes identificadas nos textos, confirma a existencia de varios problemas nas producións escritas do alumnado. Ao carón da presenza de formas non estándar no ámbito ortográfico e no léxico, moi en especial pola existencia de numerosas transferencias desde o castelán, amósanse problemas de agramaticalidade, de adecuación ao rexistro académico, así como de infrautilización dos recursos lingüísticos dispoñibles, sobre todo daqueles non coincidentes co español.

En calquera caso, as cuestións analizadas nesta obra son só unha pequena mostra das múltiples posibilidades de análise que ofrecen os textos do corpus. Con este volume, ademais de botar luz sobre os aspectos concretos seleccionados nos diferentes capítulos que o conforman, pretendemos estimular o desenvolvemento de novas investigacións a partir dos textos de *CORTEGAL*, así como a elaboración de propostas didácticas que contribúan a resolver os problemas detectados e analizados nas investigacións lingüísticas. Lémbrese que o corpus está dispoñible en aberto para calquera persoa que o desexe consultar, tal e como corresponde a calquera traballo destas características desenvolvido con financiamento público.

Non podemos pechar esta introdución sen deixar constancia do noso agradecemento a todas as persoas que contribuíron a que este volume vise a

² Véxase tamén: Álvarez de la Granja (2020a), sobre a flexión verbal nos textos de *CORTEGAL*, Álvarez de la Granja (2020b), sobre o emprego do infinitivo flexionado no mesmo corpus, e Álvarez de la Granja (2023), sobre a diferenciación entre problemas léxicos e semánticos.

luz. En primeiro lugar, e de xeito máis directo, ás autoras e autores dos capítulos que o conforman, ás persoas responsables de *Verba* e ás avaliadoras e avaliadores que revisaron a proposta inicial e contribuíron á súa mellora. En segundo lugar, ao equipo de traballo que participou na elaboración de *CORTEGAL*³, sen o cal este libro non existiría, así como ás autoras e autores dos traballos previos, amentados en §1 e na bibliografía dos diferentes capítulos. Queremos tamén mostrarlles a nosa gratitude ás entidades financiadoras tanto do volume como de *CORTEGAL*⁴ e, finalmente, ás persoas usuarias deste último e ás/aos lectoras/es do libro, porque só o emprego do corpus e a lectura da obra lle dan sentido e valor ao traballo desenvolvido.

REFERENCIAS BIBLIOGRÁFICAS

- ABEL, A. e GLAZNIEKS, A. (2017). *KoKo: Bildungssprache im Vergleich: korpusunterstützte Analyse der Sprachkompetenz bei Lernenden im deutschen Sprachraum - ein Ergebnisbericht*. Eurac Resarch 1.0. http://www.korpus-suedtirol.it/KoKo/Documents/Ergebnisse_Dokumentation_gesamt_FINAL.pdf
- ÁLVAREZ DE LA GRANJA, M. (2020a). Análise da conxugación verbal no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL). En L. de Castro Moutinho, R. L. Coimbra e A. Gómez Bautista (Coords.), *Línguas minoritarias e variação lingüística* (pp. 150-173). Universidade de Aveiro. <https://doi.org/10.34624/rj68-vz44>
- ÁLVAREZ DE LA GRANJA, M. (2020b). O infinitivo flexionado no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL). *Madrygal. Revista de estudos gallegos*, 23, 15-38. <https://doi.org/10.5209/madr.73064>
- ÁLVAREZ DE LA GRANJA, M. (2023). Os problemas léxicos e semánticos en CORTEGAL, *Corpus de textos galegos escritos por estudantes no ámbito académico*. En S. M. Torres-Outón e R. Méndez González (Eds.), *Realidades y desafíos de la tecnología aplicada a la traducción e interpretación* (pp. 169-185). Peter Lang.
- ÁLVAREZ DE LA GRANJA, M. e GONZÁLEZ SEOANE, E. (Eds.) (s.d.). *Corpus de textos galegos escritos por estudantes no ámbito académico*. Instituto da Lingua Galega. <http://ilg.usc.gal/cortegal/>

³ <http://ilg.usc.gal/cortegal/index.php?action=team>

⁴ A elaboración do corpus *CORTEGAL* foi posible grazas ao apoio financeiro da Secretaría Xeral de Política Lingüística da Xunta de Galicia a través de convenios de colaboración co Instituto da Lingua Galega e ao Ministerio de Ciencia e Innovación e Universidades (PGC2018-096069-B-I00). A maquetación do volume realizouse grazas á colaboración do Ministerio de Cultura e Deporte, mediante subvención ao Instituto da Lingua Galega.

- BATTANER ARIAS, P. e TORNER CASTELLS, S. (Eds.) (2005). *El Corpus PAAU 1992: estudios descriptivos, textos y vocabulario*. Documenta Universitaria.
- CANALE, M. e SWAIN, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47. <https://doi.org/10.1093/applin/1.1.1>
- CORDER, S. P. (1967). The Significance of Learners' Errors. *International Review of Applied Linguistics*, 5, 161-170. <https://doi.org/10.1515/iral.1967.5.1-4.161>
- DAGNEAUX, E., DENNESS, S. e GRANGER, S. (1998). Computer-aided error analysis, *System*, 26, 126-174. [https://doi.org/10.1016/S0346-251X\(98\)00001-3](https://doi.org/10.1016/S0346-251X(98)00001-3)
- DÍEZ BEDMAR, M. B. (2011). Spanish pre-university students' use of English: CEA results from the University Entrance Examination. *International Journal of English Studies*, 11(2), 141-158. <https://doi.org/10.6018/ijes/2011/2/149681>
- ESTEBAN RADÍO, S. (1997). *O vocabulario galego-castelán no ensino en Galicia: unha aproximación sociolingüística*. Galiza Editora.
- IGLESIAS RÁBADE, L. (Coord.) (1999). *Análisis de los errores del examen de inglés en las pruebas de acceso a la Universidad en el distrito universitario de Galicia*. Universidade de Santiago de Compostela.
- LOREDO, X. e SILVA VALDIVIA, B. (Coords.) (2020). *Avaliación da competencia bilingüe nos idiomas galego e castelán do alumnado de 4º da ESO*. Real Academia Galega.
- LOREDO GUTIÉRREZ, X. e SILVA VALDIVIA, B. (2014). Variables asociadas á competencia gramatical e léxica en galego e castelán do alumnado de Galicia ao final do ensino obrigatorio. *Revista de investigación en educación*, 12(2), 191-208.
- LOREDO GUTIÉRREZ, X., FERNÁNDEZ SALGADO, X. A. e SUÁREZ FERNÁNDEZ, I. (2005). Análise da expresión escrita en galego e castelán ao final do ensino obrigatorio. Resultados dun estudo realizado na comarca de Santiago de Compostela. *Cadernos de lingua*, 27, 109-126. <https://doi.org/10.32766/cdl.27.58>
- MENDES, A., ANTUNES, S., JANSSEN, M. e GONÇALVES, A. (2016). The COPLÉ2 corpus: A learner corpus for Portuguese. En N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk e S. Piperidis (Eds.), *Proceedings of the Tenth Language Resources and Evaluation Conference (LREC 2016)* (pp. 3207-3214). European Language Resources Association (ELRA). <https://aclanthology.org/L16-1511.pdf>
- MERLIN project (2014). *Annotation guidelines*. <https://merlin-platform.eu/docs/Annotation%20guidelines.pdf>
- NOIA CAMPOS, M. C. (1980). *Descrición e medida do bilingüismo en dous grupos escolares da zona de Vigo*. [Tese de doutoramento inédita]. Universidade de Santiago de Compostela.
- REZNICEK, M., LÜDELING, A., KRUMMES, C., SCHWANTUSCHKE, F., WALTER, M., SCHMIDT, K., HIRSCHMANN, H. e ANDREAS, T. (2012). *Das Falko-Handbuch: Korpusaufbau und Annotationen. Version 2.01*. Institut für deutsche

Sprache und Linguistik, Humboldt-Universität zu. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/view>

SILVA VALDIVIA, B. (2006). *Castelanismos morfosintácticos no galego. Avaliación da competencia do alumnado*. Universidade de Santiago de Compostela.

SILVA VALDIVIA, B. (Dir.) (2010). *Avaliación da competencia do alumnado de 4º da ESO nos idiomas galego e castelán*. Universidade de Santiago de Compostela.

SPILLNER, B. (2017). *Error Analysis in the World. A Bibliography*. Frank & Timme.

CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico.

Deseño do corpus e caracterización dos textos

MARÍA ÁLVAREZ DE LA GRANJA

Instituto da Lingua Galega-Universidade de Santiago de Compostela

REYES RODRÍGUEZ RODRÍGUEZ

Instituto da Lingua Galega-Universidade de Santiago de Compostela

1. INTRODUCCIÓN

CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico, é un corpus anotado de textos escritos en lingua galega por alumnado de Galicia nas probas de acceso á universidade. O obxectivo fundamental deste corpus, elaborado no Instituto da Lingua Galega da Universidade de Santiago de Compostela e accesible en <http://ilg.usc.gal/cortegal/>, é coñecer algunhas características da escrita académica do estudiantado galego ao remate da educación secundaria. Particularmente, a partir da súa análise poderán detectarse as principais eivas e carencias do alumnado na expresión escrita en lingua galega e artellar *a posteriori* estratexias e recursos didácticos focalizados nos aspectos máis problemáticos. Tamén confiamos en que a análise do corpus sirva para facilitar e promover outros estudos sobre a escrita en lingua galega do alumnado, máis alá das diverxencias do estándar e particularmente no ámbito discursivo. Así mesmo, un terceiro obxectivo é converter *CORTEGAL* nunha ferramenta didáctica de aplicación directa nas aulas, de tal xeito que o alumnado traballe con textos escritos por outros/outras estudantes, moi en especial no marco da aprendizaxe guiada por datos («data driven learning») (Gilquin e Granger 2022; McEnery e Richard 2010; Römer 2011; Timmis 2015; Vázquez Rozas e Blanco 2022).

A finalidade deste capítulo é dobre. Por un lado, pretendemos realizar unha presentación xeral da metodoloxía empregada na elaboración de *CORTEGAL*, tanto no relativo á confección da mostra como no que ten que ver coa transcripción, anotación e estandarización das redaccións que conforman o corpus. En primeiro lugar, en §2.1, describiremos os criterios de selección dos textos e ofreceremos unha caracterización xeral destes desde o punto de vista cualitativo (tipo de textos, temas, cualificacións...). A seguir, en §2.2, amosaremos, en liñas xerais, o proceso de tratamento dos textos levado a cabo na plataforma *TEITOK*, dando conta das anotacións que conteñen. A continuación, en §2.3, abordaremos con máis detalle a súa transcripción e en §2.4 a asignación de metadatos. A seguir, en §2.5 deterémonos na estandarización e codificación das formas non estándar e en §2.6 trataremos a asignación de lema e categoría gramatical. Finalmente, en §2.7 ofreceremos os trazos xerais da anotación dos conectores que serven para vincular enunciados. Esta presentación servirá como marco xeral para situar e entender adecuadamente os datos que se ofrecerán nas seguintes contribucións da obra e na segunda parte deste capítulo.

Nesta segunda parte presentaremos os datos cuantitativos globais do corpus, atendendo en §3.1 á caracterización xeral dos textos (número de palabras e lemas, media de parágrafos por texto, media de enunciados por parágrafo...), en §3.2 ao número de *tokens* e á súa distribución en clases de palabras, en §3.3 ás diferentes desviacións do estándar académico detectadas nos seis niveis analizados (ortográfico, morfolóxico, léxico, gramatical [sintáctico], semántico e discursivo) e finalmente, en §3.4, á presenza dos conectores entre enunciados. Finalmente, ofreceremos unhas conclusións en que se recollerá unha caracterización global dos textos que conforman o corpus atendendo á análise realizada previamente e unhas reflexións finais sobre *CORTEGAL*, a súa elaboración e o seu interese de cara á mellora da destreza da escritura en lingua galega.

2. METODOLOXÍA DE ELABORACIÓN DE *CORTEGAL*

2.1. Os textos e a mostra

CORTEGAL é un corpus conformado por 1000 textos manuscritos redactados no ano 2017 por estudantes de Galicia no marco da proba de acceso á universidade denominada ABAU, «Avaliación do Bacharelato para o acceso á Universidade». Máis concretamente, os textos corresponden á proba de comentario que o alumnado debe realizar no exame de Lingua Galega e Literatura. Neste exame solicítase a redacción dun texto de carácter

argumentativo, de entre 200 e 250 palabras, sobre un tema determinado, vinculado cun texto previo. As redaccións, proporcionadas pola Comisión Interuniversitaria de Galicia (CIUG) cos correspondentes permisos para a elaboración do corpus, corresponden tanto á convocatoria de xuño como á de setembro do curso 2016-2017. Presentáronse ao exame un total de 8669 estudantes en xuño e 1197 en setembro, de tal modo que a mostra de *CORTEGAL* supón un 9,87% do total de exames presentados.

En cada unha destas convocatorias propóñense dous modelos de exame diferentes dos que o estudantado debe escoller un. O corpus ofrece redaccións correspondentes ás catro opcións posibles para a pregunta de comentario, que son as que se indican a seguir:

XUÑO 2017

Opción A (texto inicial de Fran Alonso en *Dorna 27*, 2001)

Nos últimos anos a gastronomía e a cociña acadaron moita popularidade. Redacta un texto expoñendo a túa opinión sobre este fenómeno: as súas causas, o que ten de moda pasaxeira ou de cambio cultural máis duradeiro...

Opción B (texto inicial de J. Luís Sucasas en *Vieiros*, 2009)

Redacta un texto sobre a importancia que teñen o consumo e a produción (ou o consumismo e a produtividade) no noso modo de vida actual.

SETEMBRO 2017

Opción A (texto inicial de Xavier Quiroga de *Zapatillas rotas*, 2014)

Expón, de maneira argumentada, a túa opinión persoal sobre o problema que reflicte o texto e, en xeral, sobre este tipo de conflitos familiares entre pais e fillos adolescentes.

Opción B (texto inicial de Mercedes Queixas, en *Palavra Comum*, 09/10/2015)

A autora móstrase crítica co feito de que a infancia e a mocidade soñe con ser futbolista ou modelo moi maioritariamente (líña 10). Redacta un texto expoñendo de maneira argumentada o teu acordo ou desacordo co seu punto de vista.

Os centros educativos aos que pertence o alumnado que realiza as probas ABAU están asignados a 26 Comisións Delegadas (en diante CD), de tal xeito que cada CD abrangue unha zona xeográfica ampla que recolle centros tanto públicos como privados e con alumnado de procedencia diversa (urbana, periurbana, vila ou rural). A listaxe de CD cos centros educativos asignados no curso 2016-2017 pode consultarse na páxina do corpus. O número de textos da mostra de *CORTEGAL* é proporcional á cifra total de exames por CD e convocatoria (xuño e setembro), con dúas puntualizacións: por un lado, elimináronse os 29 exames correspondentes á CD 25, que recolle probas de alumnado con necesidades específicas procedente de toda Galicia e non dunha zona xeograficamente restrinxida; por outro lado, aínda que a

distribución real de exames entre xuño e setembro é de 87,9%-12,1%, na mostra de *CORTEGAL* a distribución mudouse lixeiramente (89,8%-10,2%), para reducir o peso do alumnado que realizou o exame tanto en xuño como en setembro (ao suspender na primeira convocatoria). En Álvarez de la Granja (2018, p. 58) explícase con máis detalle o procedemento seguido para levar a cabo esta modificación, que tivo en conta o número de suspensos na primeira convocatoria. A distribución exacta de exames por convocatoria e temática na mostra de *CORTEGAL* é a que se ofrece na Táboa 1:

TÁBOA 1. Distribución de exames por convocatoria e temática na mostra de *CORTEGAL*

Convocatoria	Tema	Número de textos da mostra	Porcentaxe sobre o total de textos
Xuño	A gastronomía	449	44,9%
	Consumo e produción	449	44,9%
<i>Total xuño</i>		<i>898</i>	<i>89,8%</i>
Setembro	Conflitos familiares	51	5,1%
	Os referentes da mocidade	51	5,1%
<i>Total setembro</i>		<i>102</i>	<i>10,2%</i>
<i>Total</i>		<i>1000</i>	<i>100%</i>

O reparto por temas non se fixo atendendo ás escollas do alumnado, senón que, tal e como se pode ver na táboa anterior, se procurou unha distribución equitativa entre as dúas opcións de cada convocatoria, para deste xeito asegurar a maior variedade posible nos textos e, daquela, no léxico. Máis alá desta circunstancia, a selección de textos para cada comisión foi aleatoria.

A nota media dos exames de Lingua Galega e Literatura de *CORTEGAL* é de 5,90 sobre 10 (cunha desviación estándar de 1,72) e da proba concreta de comentario de 6,75 sobre 10¹ (cunha desviación estándar de 1,93), co reparto entre xuño e setembro que se amosa na Táboa 2, onde se perciben diferenzas importantes entre as dúas convocatorias:

¹ As notas do comentario realízanse en realidade sobre 3 puntos, pero están convertidas aquí a base 10.

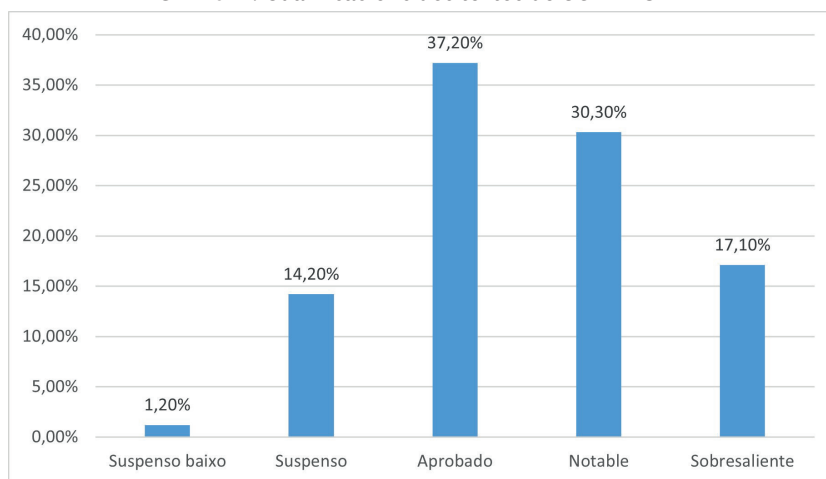
TÁBOA 2. Notas numéricas da mostra de *CORTEGAL*

Convocatoria	Exame	Desviación estándar	Comentario	Desviación estándar
<i>Xuño</i>	6,09	1,65	6,93	1,87
<i>Setembro</i>	4,26	1,43	5,21	1,75
<i>Xuño e setembro</i>	5,90	1,72	6,75	1,93

A nota media do conxunto dos exames de Lingua Galega e Literatura das probas ABAU foi de 5,77 en xuño e de 4,17 en setembro. Vemos, pois, que as medias son bastante próximas ás da mostra, cunha diferenza de 0,32 en xuño e de 0,09 en setembro, sendo sempre máis alta a media dos textos do corpus. Carecemos de datos sobre os resultados da pregunta de comentario para realizar unha comparativa similar.

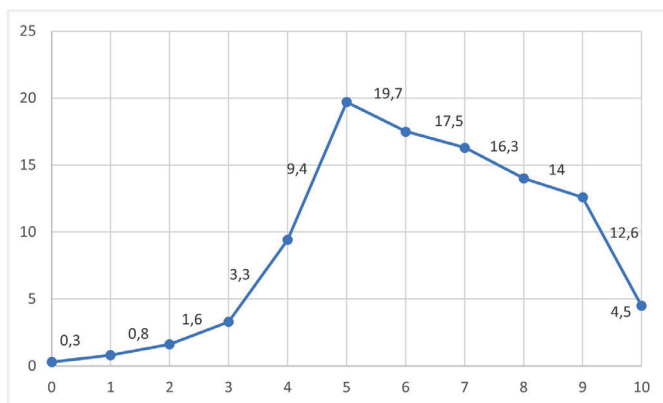
O rango de notas dos textos de *CORTEGAL* sitúase no exame de xuño entre 1 (nun único exame) e 10 (en seis exames) e no exame de setembro entre 0 e 8 nunha única proba para cada cualificación. No caso do comentario, o rango vai desde 0 nun só exame ata 10 en 45 exercicios en xuño, e dende 0 nun dos comentarios ata 9,33 noutro en setembro.

Na Gráfica 1 recóllese a distribución das cualificacións nos textos de *CORTEGAL*, onde Suspenso baixo abrangue desde 0 ata 2,4, Suspenso desde 2,5 ata 4,9, Aprobado desde 5 ata 6,9, Notable desde 7 ata 8,9 e Sobresaliente desde 9 ata 10. A cualificación máis frecuente é Aprobado, cun 37,20% dos exercicios. O 84,60% do alumnado superou a pregunta, mentres que o restante 15,40% obtivo unha cualificación inferior a 5.

GRÁFICA 1. Cualificacións dos textos de *CORTEGAL*

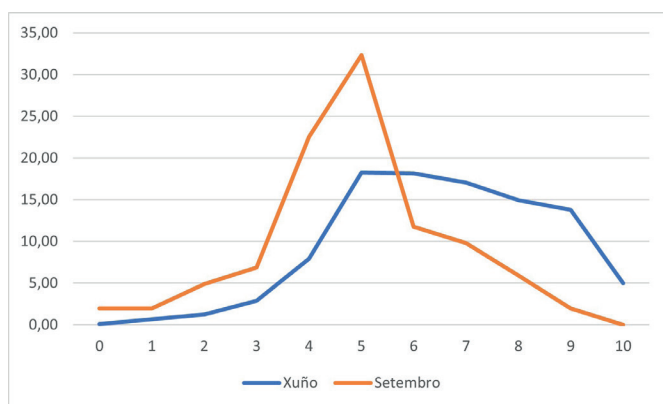
Na seguinte gráfica de dispersión (Gráfica 2), onde as cifras indican a porcentaxe de textos que obtiveron unha cualificación situada en cada rango numérico (0-0,9; 1-1,9 etc.), percíbese con máis detalle a distribución das notas. Comprobamos como o rango con maior número de exames é o situado entre 5 e 5,9, aínda que a nota concreta máis repetida, ou moda, é 6,67, moi próxima da media.

GRÁFICA 2. Distribución das cualificacións numéricas nos textos de *CORTEGAL*



Mais se se afonda en cada unha das convocatorias a través da Gráfica 3, comprobábase como en xuño se debuxa unha liña relativamente estable dende a nota 5 cara adiante, en que se concentra máis do 87% do alumnado, mentres que en setembro, do 62% de aprobados, máis da metade pertencen ao rango 5-5,9 e a partir deste, cae dous terzos a cifra de cualificacións no rango 6-6,9, seguido dun descenso considerable a medida que se incrementa a cualificación.

GRÁFICA 3. Distribución das cualificacións numéricas dos textos de *CORTEGAL* en xuño e setembro



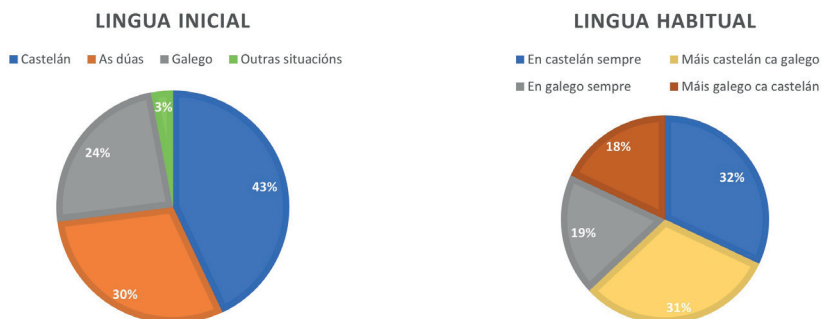
Malia estas diferenzas, compróbase como as dúas convocatorias coinciden en que a marca máis elevada se sitúa no rango 5-5,9 (aínda que en xuño a diferenza é mínima en relación co seguinte rango) e en que hai máis textos aprobados ca suspensos.

As cualificacións obtidas no exame de Lingua Galega e Literatura e na proba de comentario son uns dos poucos datos «extratextuais» que posuímos sobre as redaccións. Precisamente, un dos principais problemas implícitos no emprego de textos das probas ABAU é o feito de carecermos de datos sociolingüísticos sobre o alumnado que os redactou. Os exames son totalmente anónimos e non inclúen información de ningún tipo sobre xénero, lingua habitual, lingua inicial, tipo de residencia, nivel sociocultural etc. O único dato que posuímos corresponde, como indicamos, á CD ao que está asignado o centro educativo (este último tamén é descoñecido), de tal xeito que podemos atribuír a cada texto unha zona xeográfica ampla correspondente ao centro de ensino ao que pertence a/o estudante. Sabemos, en calquera caso, que se trata na súa maior parte de alumnado que acaba de rematar 2º de Bacharelato, maioritariamente de 17 ou 18 anos, e que ten o galego como L1 ou L2. Cómpre sinalar, de todos os xeitos, que o galego é lingua instrumental empregada na impartición de diversas materias ao longo de todos os cursos da educación primaria e da educación secundaria, así como obxecto de aprendizaxe nunha materia específica tamén en todos os cursos destes dous niveis educativos². De acordo co indicado no artigo 14 da Lei 3/1983 de Normalización Lingüística, o estudantado, ao remate do ensino obrigatorio, deberá coñecer o galego, nos niveis oral e escrito, «en igualdade co castelán».

Dado que carecemos de datos sobre a lingua inicial e habitual das persoas autoras dos textos de *CORTEGAL*, podemos presentar a información ofrecida polo Instituto Galego de Estatística (IGE) (2019) na última «Enquisa estrutural a fogares», realizada en 2018, un ano despois de que fosen escritos os textos de *CORTEGAL*. Amosamos, na Gráfica 4 e na Gráfica 5, os datos relativos á franxa de idade en que se incluírían as persoas autoras das redaccións (entre 15 e 29 anos).

² O alumnado procedente doutras comunidades ou doutros países que se incorpore ao sistema educativo galego pode solicitar a exención da materia de lingua galega durante un máximo de dous anos consecutivos (<https://www.xunta.gal/dog/Publicados/2014/20140219/AnuncioG0164-120214-0003-gl.html>). O estudandante que, por este motivo, non curse tal materia nun ou nos dous cursos de Bacharelato está exento da realización do correspondente exame nas probas ABAU (https://www.xunta.gal/dog/Publicados/2011/20110404/AnuncioEF0A_es.html; <https://www.boe.es/buscar/pdf/2016/BOE-A-2016-7337-consolidado.pdf>). Un total de 50 alumnos matriculados nas ABAU no curso 2016-2017 non realizaron o exame de Lingua Galega e Literatura.

GRÁFICAS 4 E 5. Lingua inicial e lingua habitual da mocidade galega (15-29 anos)
(Fonte: IGE, 2019)



Como se pode observar, a lingua inicial maioritaria é o castelán (43%), pero un 54% das persoas enquisadas indican que a súa lingua inicial é ben o galego, ben o galego e o castelán. Con respecto á lingua habitual, un 63% di falar só castelán ou máis castelán ca galego, mentres que un 37% indica que só fala galego ou que fala máis galego ca castelán. Evidentemente, a extrapolação destas porcentaxes a *CORTEGAL* cómpre facela con cautela, dado que a franxa de idade con que traballa o IGE (15-29 anos) é moito máis ampla ca a que corresponde, en liñas xerais, ao alumnado do corpus (17-18 anos).

No momento de seleccionar os textos das probas ABAU, eramos conscientes de que estes presentaban algunhas desvantaxes: a xa comentada imposibilidade de ter información sociolingüística sobre as persoas autoras das redaccións, a necesidade de transcribilas manualmente e a tensión derivada da importancia da proba e das restricións de tempo con que se elaboraron estas, que determinaron probablemente que o resultado non fose o mellor dos posibles (González Álvarez, 1990, p. 209). Con todo, o feito de traballar con textos das probas ABAU ofrece tamén varias vantaxes, tal e como se indica en Álvarez de la Granja (2018, p. 57). Entre elas, a dispoñibilidade inmediata dos textos, a súa homoxeneidade, así como a garantía de que a/o estudante escribe a redacción con seriedade, procurando utilizar sempre a variedade estándar (posto que calquera diverxencia pode ser obxecto de sanción na cualificación). Ademais, o feito de traballar con textos manuscritos permite, como veremos en §2.3, ofrecer información sobre as formas riscadas e as engadidas *a posteriori*, co que resulta posible rastrexar o proceso compositivo do alumnado. Finalmente, e de maneira especial, os textos das ABAU, pola súa transcendencia, teñen un valor engadido que dota o corpus, e os resultados que se extraían da súa análise, dunha especial relevancia no marco da sociedade galega.

O conxunto dos 1000 textos é accesible a través da pestana «Textos» da páxina do corpus (<http://ilg.usc.gal/cortegal/index.php?action=texts>). De igual xeito, se non se escribe nada na caixa do buscador e se preme en «Buscar» obtense tamén a listaxe de todas as redaccións, desta volta cos seus metadatos (*vid.* §2.4).

2.2. Tratamento dos textos en TEITOK

Os textos de *CORTEGAL* están transcritos, *tokenizados* e anotados na plataforma *TEITOK*, unha plataforma utilizada na creación de diferentes corpus, especialmente, aínda que non só, corpus históricos e de aprendentes, e de gran interese para o tratamento de textos manuscritos pola posibilidade de combinar anotacións lingüísticas (lema, categoría morfosintáctica...) e textuais (fragmentos borrados, cursivas...) (Janssen, 2016, p. 4037)³.

TEITOK is a framework for creating, maintaining, and publishing annotated corpora. It is a web-based environment written mostly in a combination of PHP and Javascript. In TEITOK, a corpus consists of a collection of XML files, each in the Text Encoding Initiative (TEI) format (...), with a slightly modified tokenization system (...). The system makes it easy to display each XML file (...), edit metadata (...) and individual tokens (...), and search through the corpus (...) (Janssen, 2016, p. 4037).

O proceso de traballo seguido cos diferentes textos que conforman o corpus *CORTEGAL* é o que se indica a seguir. Todas as tarefas, agás a primeira, foron realizadas en *TEITOK*:

- 1) Escaneado dos 1000 textos.
- 2) Creación dun arquivo XML coa transcripción manual do texto e etiquetas TEI que identifican a súa estrutura e algúns elementos concretos dos orixinais (formas riscadas, engadidas *a posteriori*...) (*vid.* §2.3).
- 3) Revisión das transcripcións.
- 4) Introducción dos datos fixos da cabeceira e dalgúns datos variables (título, CD, cualificacións, convocatoria, tema, responsables da transcripción e revisión) (*vid.* §2.4).
- 5) *Tokenización*. Cada *token*, incluíndo signos de puntuación, é representado por un elemento <tok>. Debe terse en conta a este respecto que, no caso das contraccións e das unidades formadas por verbo e pronomes enclíticos, se emprega un sistema mixto, de tal xeito que toda a unidade é

³ En <http://teitok.corpuswiki.org/> pódese atopar unha listaxe de proxectos elaborados nesta plataforma, así como unha guía para o seu emprego.

un elemento <tok> conformado por varios *dtokens* (<dtok>) (*vid. infra* exemplo 2):

One of the eternal problems in annotated corpora is how to handle contractions: whether to treat them as one token or two. That is why TEITOK takes a mixed approach: the <tok> elements are roughly speaking orthographic words. In the case of contractions, two or more grammatical words are inserted as children of the <tok>, called <dtok>. In this manner, it is possible to associate POS and lemma to the grammatical words, but associate the normalized orthography to the orthographic word (Janssen, 2016, p. 4038).

- 6) Estandarización e codificación das formas non estándar en seis niveis lingüísticos (*vid. §2.5*).
- 7) Lematización automática mediante *FreeLing*, con asignación de lema e categoría gramatical, e revisión manual posterior (*vid. §2.6*).
- 8) Asignación manual de lemas e categorías gramaticais orixinais (*vid. §2.6*).
- 9) Anotación dos conectores que vinculan enunciados (*vid. §2.7*).
- 10) Introducción da información restante na cabeceira dos textos (datos cuantitativos sobre os textos e responsables do proceso de anotación e lematización) (*vid. §2.4*).
- 11) Revisión da estandarización e codificación.
- 12) Revisión xeral do corpus.

No arquivo XML creado para cada texto, as anotacións lingüísticas son representadas normalmente como atributos dos diferentes *tokens* (para excepcións, *vid. §2.5.4*). No exemplo (1) que ofrecemos a seguir, a forma escrita polo estudante, *recetas* (*token* 137 do texto), recibe as seguintes anotacións: a forma lexicamente estandarizada, *receitas*, o lema orixinal *receta*, a identificación do tipo de desviación do estándar mediante o código L_w_su, a orixe desta desviación, a través do código L_sp, o lema estándar *receita* así como a súa categoría gramatical NCFP000.

(1) <tok id=»w-137» form=»recetas» lform=»receitas» olemma=»receta»
 problem=»L_w_su» psource=»L_sp» lemma=»receita»
 pos=»NCFP000»>recetas</tok>

No exemplo (2) pode verse a división en <dtok> que se leva a cabo nas contraccións e a asignación aos <dtok> do lema e categoría gramatical:

(2) <tok id=»w-114»>no<dtok lemma=»en» pos=»SP» id=»d-114-1» form=»en»/><dtok lemma=»o» pos=»DAoMSO» id=»d-114-2» form=»o»/></tok>

Á hora de visualizar os textos, todas as anotacións lingüísticas asociadas a un *token* son visibles ao colocar o cursor sobre este, tal e como se pode comprobar na Figura 1, en que nos situamos sobre *vivenzas*:

FIGURA 1. Visualización das anotacións lingüísticas dos *tokens* (I)

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico Estándar léxico Estándar gramatical
 Estándar semántico Estándar discursivo

Mostrar: **Cores** **Alliñación** <pb> <lb>

Anotación: Lema estándar Lema ordinal Clase de palabra (estándar) Clase de palabra (ordinal)
 Tipo de desviación do estándar Fonte da forma non estándar Corrección derivada Conector

Estamos ante un texto da autoría de J. Luís Sucasas publicado en "Viveiros". Como tema principal trátase a actividade da produción e do consumo así como as distintas **vivenzas** e interpretacións en cada caso.

É sabido que, non hai consumo sen produción. Ámbolos dous son esixibles. Hoxe en día lévanse actividades de produción para satisfacer as necesidades do ser humano. Conforme avanza os : necesidades xa que todos sabemos que se nos cumpren os caprichos á primeira, non resultan t se andamos detrás deles unha tempada. Isto é o que precisamente se está perdendo hoxe en día : innovación exerce un exceso de materias de posible adquisición. Por outra banda, os humanos : sempre [] o último e o mellor de todo. Precisamente este tema é o que marca as diferenzas de : temos a oportunidade de adquirir as mesmas cousas e isto pode causar un enfrontamento. Exis : ante isto que precisamente aparecen remarcadas no texto. Todo é unha cadea, a muller que aco : contra do baixo prezo da leite ve ameazada o seu poder adquisitivo e actúa ante o problema xa : **Porén**, a empregada de Toyopes ao enterarse da súa diminución no salario recorre a comprar leit : maneira reducir gastos. Pero é precisamente este feito o que fai que se consuma o leite máis ba :
 O apuro é cego e a fame actúa sen mirar, cada un mira polo seu e non polo ben común. Nisto se basa a sociedade de : hoxe en día: O mundo está mal repartido, así como a fame no mundo.

vivenzas	
Estándar léxico	vivenzas
Lema estándar	vivenza
Lema orixinal	vivenza
Clase de palabra (estándar)	Nome (NCFP000) Common; feminine; plural
Tipo de desviación do estándar	L_w_su
Fonte da forma non estándar	L_anal,L_hc

Ademais, premendo nos botóns das diferentes anotacións (*vid.* Anotación en Figura 1 e 2) estas fanse visibles baixo os *tokens*. Na Figura 2, premeuse en lema estándar e clase de palabras (estándar) no texto anterior (ofrécese só o primeiro parágrafo):

FIGURA 2. Visualización das anotacións lingüísticas dos tokens (II)

Opcións de visualización

Texto: **Transcripción completa** | Versión final estudante | Estándar ortográfico | Estándar léxico | Estándar gramatical
 Estándar semántico | Estándar discursivo

Mostrar: **Cores** | Aliñación | <pb> | <lb>

Anotación: Lema estándar | Lema ordinal | Clase de palabra (estándar) | Clase de palabra (ordinal)
 Tipo de desviación do estándar | Fonte da forma non estándar | Corrección derivada | Conector

Estamos	ante	un	texto	da	autoría	de	J.	Luis	Sucasas	publicado	en	"	Viveiros
Verbo	Preposición	Determinante	Nome	Preposición+Determinante	Nome	Preposición	Nome	Nome	Nome	Verbo	Preposición	Puntuación	Nome
estar	ante	un	texto	de+o	autoría	de	José	Luis	Sucasas	publicar	en	"	viveiro
"	.	Como	tema	principal	trátase	a	actividade	da	produción	e	do		
Puntuación	Puntuación	Conxunción	tema	Adxectivo	Verbo+Pronome	Determinante	Nome	Preposición+Determinante	Nome	Conxunción	Preposición+Determinante		
"	.	como	tema	principal	tratar+se	o	actividade	de+o	produción	e	de+o		
consumo	así	como	as	distintas	vivencias	e	interpretacións	en	cada	caso	.		
Nome	Adverbio	Conxunción	Determinante	Adxectivo	Nome	Conxunción	Nome	Preposición	Determinante	Nome	Puntuación		
consumo	así	como	o	distinto	vivencia	e	interpretación	en	cada	caso	.		

2.3. Transcripción dos textos

No proceso de transcripción dos textos reproducimos fielmente o manuscrito orixinal e, deste xeito, mantemos as grañas, a puntuación, a acentuación, as abreviaturas e a unión e separación de palabras da/do estudante. Ademais, respectamos a construción de cada texto, reproducindo a súa organización en parágrafos, a súa distribución en páxinas (no caso de que as redaccións se estendan por máis dunha), así como o inicio e o remate de cada liña.

Todas as redaccións son transcritas entre as etiquetas <text> e </text>. Ao longo do texto márcase o comezo de cada liña mediante <lb/> e o remate e inicio dos distintos parágrafos que conforman o texto mediante <p> e </p>. Ademais, tamén se sinalan as páxinas a través da etiqueta <pb/>, que se inclúe aínda que haxa unha soa páxina. Ofrecemos a modo de exemplo un texto transcrito con estas e outras etiquetas, que se comentarán nesta epígrafe:

- (3) <text><p><pb/><lb/>A gastronomía e a cociña comezaron a ter importancia, dende <lb/>sempre, xa que comer foi unha cousa moi importante toda a vida, <lb/>pero co paso do tempo a xente comezou a cambiar de formas <lb/>de pensar, e empezousell<unclear>a</unclear> empezouselle a dar máis importancia <lb/>ao estético, e eu creo que dende eso comezou o cambio na <lb/>gastronomía tamén. Para querer unha cousa primeiro ten que «en<lb/>trar polos ollos», polo que cada paso a cociña comezou a ser máis <lb/>estética, ata tal punto que, a día de hoxe, é máis importante <pb/><lb/>que quede ben no prato a que quite a fame.</p> <p><lb/>Hai novas técnicas como a deconstrución, que consiste en se<lb/>parar os elementos dunha comida e cocíalos por separado. <lb/>Tamén se fan concursos de cociña, nos que non só se valora a presen<lb/>tación, tamén

o sabor e que o prato esté ben cociñado.

Eu creo que todo eso está ben, innovar e que quede ben no prato é algo importante, pero creo que o máis importante é que a comida serve para comer, polo que ten que quitar a fame, e non é un xoguete co que se poida probar cousas novas, polo que penso que a cociña moderna ten o seu punto positivo ata sempre que non sexa para un fin lúdico.

A introdución das etiquetas comentadas permite visualizar o texto na súa forma orixinal, aínda que esta non é a única posibilidade existente, posto que, premendo no botón «Aliñación» das diferentes «Opcións de visualización», a vista do texto pode modificarse. Na Figura 3 vemos o texto de (3) coas liñas dispostas segundo o orixinal e coa indicación do inicio dunha nova páxina mediante unha liña horizontal; na Figura 4, as liñas discorren todo ao longo da pantalla e sen indicación do cambio de páxina e, finalmente, na Figura 5, onde está activada esta última visualización e as opcións <pb> e <lb> de «Mostrar», amósanse barriñas verticais que identifican o inicio e o final de cada liña e un número ([1]) que sinala o remate da páxina e o inicio da seguinte.

FIGURA 3. Visualización do texto na súa forma orixinal

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico Estándar morfolóxico Estándar gramatical Estándar discursivo

Mostrar: **Cores** **Aliñación** <pb> <lb>

Anotación: **Lema estándar** Clase de palabra (estándar) Tipo de desviación do estándar Fonte de forma non estándar

Corrección derivada Conector

A gastronomía e a cociña comezaron a ter importancia, dende sempre, xa que comer foi unha cousa moi importante toda a vida, pero co paso do tempo a xente comezou a cambiar de formas de pensar, e empezouselle a dar máis importancia ao estético, e eu creo que dende eso comezou o cambio na gastronomía tamén. Para querer unha cousa primeiro ten que "en trar polos ollos", polo que cada paso a cociña comezou a ser máis estética, ata tal punto que, a día de hoxe, é máis importante

que quede ben no prato a que quite a fame.

Hai novas técnicas como a deconstrucción, que consiste en se parar os elementos dunha comida e cociñalos por separado. Tamén se fan concursos de cociña, nos que non só se valora a presentación, tamén o sabor e que o prato esté ben cociñado.

Eu creo que todo eso está ben, innovar e que quede ben no prato é algo importante, pero creo que o máis importante é que a comida serve para comer, polo que ten que quitar a fame, e non é un xoguete co que se poida probar cousas novas, polo que penso que a cociña moderna ten o seu punto positivo ata sempre que non sexa para un fin lúdico.

FIGURA 4. Visualización do texto sen a disposición de liñas orixinal e sen distribución en páxinas

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico Estándar morfolóxico Estándar gramatical
 Estándar discursivo

Mostrar: **Colores** **Afiliación** <pb> <lb>

Anotación: Lema estándar Clase de palabra (estándar) Tipo de desviación do estándar Fonte de forma non estándar
 Corrección derivada Conector

A gastronomía e a cociña comezaron a ter importancia; dende sempre, xa que comer foi unha cousa moi importante toda a vida, pero co paso do tempo a xente comezou a cambiar de formas de pensar, e empezouse a dar máis importancia ao estético, e eu creo que dende eso comezou o cambio na gastronomía tamén. Para querer unha cousa primeiro ten que "entrar polos ollos", polo que cada paso a cociña comezou a ser máis estética, ata tal punto que, a día de hoxe, é máis importante que quede ben no prato a que quite a fame.

Hai novas técnicas como a deconstrucción, que consiste en separar os elementos dunha comida e cociñalos por separado. Tamén se fan concursos de cociña, nos que non só se valora a presentación, tamén o sabor e que o prato esté ben cociñado.

Eu creo que todo eso está ben, innovar e que quede ben no prato é algo importante, pero creo que o máis importante é que a comida serve para comer, polo que ten que quitar a fame, e non é un xogueito co que se poida probar cousas novas, polo que penso que a cociña moderna ten o seu punto positivo ata que sempre que non sexa para un fin lúdico.

FIGURA 5. Visualización do texto con marcas de inicio e remate de cada liña e páxina

Opciones de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico Estándar morfolóxico Estándar gramatical
 Estándar discursivo

Mostrar: **Colores** **Alineación** <pb> <lb>

Etiquetas: Lema estándar Clase de palabra (estándar) Tipo de desviación del estándar Fuente de la forma no estándar
 Corrección derivada Conector

[A gastronomía e a cociña comezaron a ter importancia; dende sempre, xa que comer foi unha cousa moi importante toda a vida, pero co paso do tempo a xente comezou a cambiar de formas de pensar, e empezouse a dar máis importancia ao estético, e eu creo que dende eso comezou o cambio na gastronomía tamén. Para querer unha cousa primeiro ten que "entrar polos ollos", polo que cada paso a cociña comezou a ser máis estética, ata tal punto que, a día de hoxe, é máis importante que quede ben no prato a que quite a fame.

[Hai novas técnicas como a deconstrucción, que consiste en separar os elementos dunha comida e cociñalos por separado. Tamén se fan concursos de cociña, nos que non só se valora a presentación, tamén o sabor e que o prato esté ben cociñado.

[Eu creo que todo eso está ben, innovar e que quede ben no prato é algo importante, pero creo que o máis importante é que a comida serve para comer, polo que ten que quitar a fame, e non é un xogueito co que se poida probar cousas novas, polo que penso que a cociña moderna ten o seu punto positivo ata que sempre que non sexa para un fin lúdico.

No primeiro tipo de visualización tamén se amosan os cortes de palabra en final de liña, tal e como se pode comprobar en varios *tokens* da Figura 3.

O respecto ao manuscrito de partida tamén se manifesta no rexistro dos engadidos, as riscaduras e as emendas que se aprecian no texto orixinal e que pertencen ao alumno ou alumna, o que, por outro lado, permite rastrexar o seu proceso compositivo. Os fragmentos borrados no texto orixinal mediante calquera sistema (riscado, borrancho etc.) márcanse mediante o elemento . Se o texto borrado é lexible, este incorpórase tras o citado elemento, como na palabra *pero*, riscada neste fragmento: «máis <lb/>»sabios», pero porque non se dan conta de que é unha profesión

<lb/>coma calquer outra». Escollendo dentro das «Opcións de visualización» o modo «Transcrición completa», os fragmentos eliminados amósanse en gris e riscados, tal e como se pode ver na Figura 6, onde se recollen varias palabras ou secuencias borradas, entre elas a forma *pero* que acabamos de comentar (última liña). Esta visualización responde á establecida na folla de estilo do corpus mediante linguaxe CSS:

FIGURA 6. Visualización das formas riscadas

Opcións de visualización

Texto: **Transcrición completa** Versión final estudante Estándar ortográfico Estándar morfolóxico Estándar léxico
 Estándar gramatical Estándar semántico Estándar discursivo

Mostrar: **Cores** Aliación <pb> <lb>

Anotación: Lema estándar Lema ordinal Clase de palabra (estándar) Tipo de desviación do estándar
 Fonte da forma non estándar Conector

Últimamente, o mundo da gastronomía está máis vixente que no pasado.

Creo que esto sucedeu polo feito de que algúns artistas decidiron empezar a utilizar ~~aos~~ pratos de comida como medio de expresión. Agora é habitual ver na televisión concursos gastronómicos, nos que o que de verdad importa é a estética e non o sabor. ~~El~~ Isto está ben pero para ~~solo~~ ese tipo de situacións, ou para quen esté disposto a facelo.

Non me parece un disparate o que se gasta algunha xente nalgúns restaurantes "modernos", ~~per~~ non porque se van a ese tipo de sitios, e para disfrutar da estética, da pequena obra de arte. Eles son libres de gastar o que queiran, igual que o son as persoas que compran calquer tipo de obra artística.

Isto non quita que haxa xente que o faga por aparentar e sentirse superior aos demais, pero ~~eiso~~ ocorre en todos os hámbitos.

Penso que a gastronomía gañou tanta importancia porque ~~me~~ mestura o mundo da arte con algo tan común como comer. En esta acción, poderíamos falar tamén da metáfora de comer arte.

Moitas persoas pensan que por traballar con algo relacionado coa arte, ~~sabeo~~ ~~todo~~ e teñen máis coñecementos ou son máis "sabios", ~~pero~~ porque non se dan conta de que é unha profesión coma calquer outra.

Premendo, dentro das «Opcións de visualización», en «Versión final estudante», as formas riscadas desaparecen e o texto pode verse na súa forma definitiva, tal e como se pode comprobar na Figura 7:

FIGURA 7. Visualización da versión final da/do estudante nun texto con formas riscadas

Opcións de visualización

Texto: **Transcrición completa** **Versión final estudante** Estándar ortográfico Estándar morfolóxico Estándar léxico
 Estándar gramatical Estándar semántico Estándar discursivo

Mostrar: **Cores** Aliación <pb> <lb>

Anotación: Lema estándar Lema ordinal Clase de palabra (estándar) Tipo de desviación do estándar
 Fonte da forma non estándar Conector

Últimamente, o mundo da gastronomía está máis vixente que no pasado.

Creo que esto sucedeu polo feito de que algúns artistas decidiron empezar a utilizar os pratos de comida como medio de expresión. Agora é habitual ver na televisión concursos gastronómicos, nos que o que de verdad importa é a estética e non o sabor. Isto está ben pero para ese tipo de situacións, ou para quen esté disposto a facelo.

Non me parece un disparate o que se gasta algunha xente nalgúns restaurantes "modernos", porque se van a ese tipo de sitios, e para disfrutar da estética, da pequena obra de arte. Eles son libres de gastar o que queiran, igual que o son as persoas que compran calquer tipo de obra artística.

Isto non quita que haxa xente que o faga por aparentar e sentirse superior aos demais, pero iso ocorre en todos os ámbitos.

Penso que a gastronomía gañou tanta importancia porque mestura o mundo da arte con algo tan común como comer. En esta acción, poderíamos falar tamén da metáfora de comer arte.

Moitas persoas pensan que por traballar con algo relacionado coa arte, teñen máis coñecementos ou son máis "sabios", porque non se dan conta de que é unha profesión coma calquer outra.

Se o texto riscado é ilexible (sexa unha palabra, un carácter, unha secuencia de palabras ou de caracteres), emprégase un elemento e no seu interior un elemento baleiro <gap/>, con indicación, se é posible, do número de caracteres («character») ou palabras («word») riscadas (mediante os atributos @quantity e @unit): «<gap quantity=»4» unit=»character»/>». Na visualización, esta etiquetaxe dá lugar a [...], tal e como se pode ver na última liña da Figura 3.

Como indicamos, tamén deixamos constancia das formas engadidas despois dunha primeira versión do texto (recoñecibles, por exemplo, porque figuran na interliña, porque se engaden mediante unha nota ou sobre líquido corrector). Etiquétanse sempre mediante un elemento <add>, con independencia do sistema que a/o estudante empregase para introducir a adición: «o resto de produtos se mercan para un <add>mesmo</add> e coa nosa actitude de derroche egoísta». O texto engadido aparece destacado en vermello, tanto na «Transcrición completa», tal e como se pode comprobar na Figura 8, como na «Versión final». Con todo, no caso de que a adición sexa un fragmento dunha palabra, o destacado en vermello só se visualiza na «Transcrición completa».

FIGURA 8. Visualización da palabra engadida *mesmo*

o resto de produtos se mercan para un **mesmo** e coa nosa actitude ~~de derroche~~ egoísta

Finalmente, tamén se etiquetan as palabras cunha lectura dubidosa (estean riscadas ou non) mediante o elemento <unclear>. O usuario visualízalas cun fondo verde azulado, tal e como se pode comprobar na Figura 9: «tendas de roupa locais, que pecharon porque non poden competir, como <unclear>-tamén</unclear> é o culpable da».

FIGURA 9. Visualización da palabra de lectura dubidosa *tamén*

tendas de roupa locais, que pecharon porque non poden competir, como **tamén** é o culpable da

Para máis detalles sobre o proceso de transcrición dos textos (por exemplo, sobre a combinación de diferentes etiquetas), poden consultarse as instrucións empregadas no proxecto e que figuran na páxina web do corpus (<http://ilg.usc.gal/cortegal/downloads/manual-transcripcion.pdf>).

2.4. Asignación de metadatos

Os metadatos proporcionados en *CORTEGAL* son de tres tipos, atendendo á clasificación proposta por Burnard (2005): metadatos analíticos, descritivos e administrativos. Distribuímoslos, empregando etiquetas TEI, en tres elementos: <fileDesc>, <encodingDesc> e <profileDesc>. Baixo o primeiro elemento, incluímos, por unha banda, datos xerais sobre o corpus: o seu nome, a indicación das entidades que o financiaron e información sobre a súa publicación (persoas e entidade responsables do corpus, lugar de publicación, ano, licenza de distribución), así como sobre a orixe dos textos e o seu carácter anónimo. Por outra banda, incluímos tamén información concreta sobre cada texto: o título de cada un deles, seguindo a estrutura ABAU/2016-2017/CDO4/xuño/03, onde o número da CD, a convocatoria (xuño ou setembro) e o número do texto dentro de cada CD son variables; as diferentes persoas responsables da transcripción, da revisión da transcripción, da lematización e da anotación lingüística do texto; e finalmente, os seguintes datos cuantitativos sobre este: número de palabras, número de lemas, densidade léxica (que resulta de dividir o número de lemas entre o de palabras), número de enunciados, media de palabras por enunciado, número de palabras do enunciado máis longo, número de palabras do enunciado máis curto, número de parágrafos e media de enunciados por parágrafo.

Para a obtención da maior parte destes últimos datos acudimos á ferramenta *DContado* (Gómez Guinovart e Solla, 2017-2022), á que subimos a versión lexicamente estandarizada dos diferentes textos de *CORTEGAL* para a súa análise cuantitativa. Debe sinalarse que o feito de analizar a versión normalizada no nivel léxico⁴, versión necesaria para levar a cabo a lematización e o cómputo de lemas en *DContado*, pode dar lugar a pequenas discrepancias entre o número de palabras do texto orixinal e o número de palabras computadas (por exemplo, cando unha unidade complexa, como *sin embargo*, se estandariza no nivel léxico cunha soa palabra, como *porén*). Por outro lado, en *Dcontado* compútanse como unha soa palabra as contraccións e as combinacións de verbo e clítico, fronte á contaxe individualizada de cada compoñente que realiza *TEITOK*, e ademais, fronte a este, non contabiliza os signos de puntuación, de xeito que o total de palabras computado por esta ferramenta é inferior ao número de *tokens* que mostra a aplicación de *CORTEGAL*. Finalmente, debe terse en conta que a análise en lemas que realiza *DContado* ofrece bastantes problemas no tratamento das contraccións e dos

⁴ A versión estandarizada no nivel léxico implica estandarización ortográfica, morfolóxica e léxica. Vid. §2.5.3.

grupos de verbo e pronome enclítico, de tal xeito que os datos que se ofrecen sobre lemas (e sobre densidade léxica) deben tomarse con cautela e son meramente orientativos.

En segundo lugar, en <encodingDesc> describimos o proceso de transcripción e codificación dos textos e, finalmente, en <profileDesc> informamos doutras características das redaccións, algunhas xerais para todas elas (ano de creación, tipo e fonte dos textos, materia á que pertencen, curso das/dos estudantes) e outras variables, como son a convocatoria, as cualificacións recibidas na pregunta de comentario e no exame de Lingua Galega e Literatura (cualificacións numéricas e cualitativas, neste último caso coas opcións e coas franxas numéricas indicadas en §2.1), o tema dos textos e a CD a que pertencen.

Tras deseñar en *TEITOK* o modelo da cabeceira, a introdución dos metadatos realízase de maneira moi sinxela a través dun formulario cos distintos campos establecidos, nos que se incorpora a información correspondente (pode fixarse un valor fixo para os campos que se desexe). Na Figura 10 ofrécese un fragmento deste formulario:

FIGURA 10. Fragmento do formulario para a introdución de metadatos

Template: *teiHeader-edit.tpl*

Título	ABAU/2016-2017/CD04/xuño/03
Ano de redacción	2017
Curso	2º de Bacharelato
Comisión Delegada	03
Convocatoria de exame	Xuño
Xénero	Argumentativo
Materia	Lingua e literatura galegas
Tema	A gastronomía

Os datos así introducidos incorpóranse automaticamente na cabeceira de cada arquivo (<teiHeader>), a cal precede o texto (<text>).

No que respecta á visualización dos metadatos, só son visibles para as persoas usuarias do corpus os que se recollen na Figura 11. Por defecto, ademais do título, unicamente se ven a CD, a convocatoria do exame e o tema, pero cun simple «clic» desprégase o resto da información.

FIGURA 11. Metadatos visibles para as persoas usuarias do corpus

ABAU/2016-2017/CD04/xuño/03

Ano de redacción	2017
Curso	2º de Bacharelato
Comisión Delegada	03
Convocatoria de exame	Xuño
Xénero	Argumentativo
Materia	Lingua e literatura galegas
Tema	A gastronomía
Número de palabras	195
Número de lemas	103
Densidade léxica	53
Número de enunciados	8
Media de palabras por enunciado	24.38
Número de palabras do enunciado máis longo	45
Número de palabras do enunciado máis curto	5
Número de parágrafos	3
Media de enunciados por parágrafo	2.67

No buscador, os metadatos que funcionan como posibles elementos de filtraxe dos textos son o título, o tema, a convocatoria e a CD, así como os datos cuantitativos sobre o texto. As cualificacións só son accesibles ás persoas con permisos de edición. Se se dispón de tales permisos, poden consultarse na cabeceira de cada texto e poden realizarse filtraxes por cualificación a través do buscador.

2.5. Estandarización e codificación das formas non estándar

Neste apartado, tras presentar o marco metodolóxico xeral en que se realiza a estandarización e etiquetaxe das formas non estándar en *CORTEGAL* (§2.5.1) e tras xustificar a realización destas dúas tarefas nos textos do corpus (§2.5.2), ofrecemos unha caracterización do sistema multinivel de estandarización (§2.5.3), mostramos as clasificacións das formas non estándar propostas e os correspondentes sistemas de codificación (§2.5.4), así como algunhas outras anotacións vinculadas coas dúas tarefas indicadas (§2.5.5). Finalmente, presentamos algúns criterios metodolóxicos empregados para minimizar a subxectividade implícita no proceso de estandarización e codificación das formas non estándar e lograr a maior homoxeneidade posible (§2.5.6).

2.5.1. Marco metodolóxico xeral

CORTEGAL sitúase na liña metodolóxica dos xa numerosos corpus de aprendentes elaborados para outras linguas (pode verse unha listaxe dalgúns destes corpus en Centre for English Corpus Linguistics, 2022). Agora ben, a maior parte dos corpus de aprendentes inclúen textos elaborados por estudantes dunha lingua estranxeira ou dunha segunda lingua, e incluso podemos encontrar esta característica como trazo definidor do concepto «corpus de aprendentes»: «[...] learner corpora, which can be defined as electronic collections of natural or near-natural data *produced by foreign or second language (L2) learners* and assembled according to explicit design criteria» (Granger et al., 2015, p. 1; a cursiva é nosa). Porén, de acordo co indicado en §2.1, os textos de *CORTEGAL* foron elaborados por estudantes que teñen o galego como lingua inicial (L1) ou como segunda lingua (L2), entendida esta última como lingua non aprendida no ámbito familiar, pero si en contexto natural e formal desde os primeiros anos de formación académica.

Con todo, cremos que é pertinente considerar *CORTEGAL* como un corpus de aprendentes, posto que o elevado número de formas non estándar atopadas nos textos (*vid.* §3.3) é boa mostra de que o alumnado non domina a destreza da escritura na variedade estándar da lingua galega e esta está aínda en proceso de aprendizaxe. Concordamos, pois, con Abel et al. (2014) na pertinencia de utilizar o termo «aprendentes» tamén para corpus que conteñen textos en L1, como sucede co corpus *KoKo* de textos en alemán (Abel et al., 2012):

We refer to people as L1 learners when they are still in the process of learning their L1 or related skills of importance such as writing and text production. (...). From a linguistic point of view, the texts written by L1 language learners are likely to have many features of non-standard writing in common with L2/FL learners. However, since some features are specific to either L1 or L2/FL learners, both learner types relate to separate learner varieties. From the perspective of computational processing, L1 and L2/FL learner corpora are fully equivalent since both are compilations of textual data that may deviate from the standard variety (Abel et al., 2014, p. 2414).

Máis en concreto, o noso corpus sitúase no marco metodolóxico coñecido como «Análise informatizada de erros» ou «Análise de erros asistida por ordenador» (Dagneaux et al., 1998), que ten como un dos seus elementos centrais a estandarización e codificación das formas ou secuencias non estándar, aspecto que trataremos nas seguintes subepígrafes de §2.5.

2.5.2. *Dous procesos: estandarización e codificación*

Tal e como indican Díaz-Negrillo e Fernández-Domínguez (2006, p. 86), nos corpus de aprendentes con análise informatizada de erros é habitual combinar dous procesos complementarios: a introdución de etiquetas identificadoras de tipos de formas desviantes⁵ e a normalización ou asignación das formas estándar correspondentes, denominadas frecuentemente *target hypothesis* (Reznicek et al., 2013, p. 104, Stemle et al., 2019, p. 428). Malia certas reticencias cara á normalización pola imposibilidade por parte do/da investigador(a) de coñecer as intencións da persoa que escribe os textos (Alonso-Ramos, 2016, p. 8), como sinalan Lüdeling e Hirschmann (2015, p. 141), «an error-annotated corpus which does not provide target hypotheses hides an essential step of the analysis», pois a codificación de cada forma non estándar está condicionada e xustificada, cando menos en parte, pola forma estándar asignada en cada caso. De acordo con este criterio, en *CORTEGAL* levamos a cabo tanto a etiquetaxe das formas non estándar como, na maior parte dos casos⁶, a súa normalización ou estandarización, tal e como se pode comprobar na Figura 12, en que se corrixe *durara* por *durará* (en «Estándar ortográfico») e ao mesmo tempo se asigna un código identificador do tipo de desviación do estándar que corresponde, unha omisión do acento gráfico (O_ac_om [Ortography_accent_omission]). En *CORTEGAL*, os procesos de estandarización e codificación realízanse simultaneamente e non en fases sucesivas.

⁵ Aínda que en certos casos, como por exemplo nas desviacións ortográficas, o emprego da palabra *erro* non supón ningún problema, cando falamos do conxunto de formas anotadas en *CORTEGAL* preferimos empregar formas non estándar ou formas diverxentes/desviantes do estándar ou da norma, en vez de usar o termo *erros*, palabra que, polas súas connotacións, resulta pouco apropiada para a etiquetaxe, por exemplo, dos dialectalismos. Incluímos ademais etiquetas para identificar voces ou construcións que se desvían do estándar académico polo seu rexistro, pero que son perfectamente correctas en niveis de lingua coloquiais e tamén neste caso o termo *erro* parece pouco afortunado. Tal e como sinalan Alonso-Ramos (2016, p. 8) ou Stemle et al. (2019, p. 15) xa noutros casos se ten evitado o emprego desta palabra, substituída, por exemplo, por *uso non convencional*, *desviación da norma* etc.

⁶ Unha das etiquetas que utilizamos informa da inintelixibilidade de todo un enunciado (*vid.* §2.5.4). Nestes casos, non ofrecemos forma normalizada. Tampouco o facemos naqueles casos en que é evidente que se está utilizando unha palabra cun significado que non lle corresponde, pero no que nos resulta imposible determinar que contido se quere transmitir.

FIGURA 12. Anotación e normalización da forma *durara* nun texto de *CORTEGAL*

durara moito máis, senon que co paso do tempo e ao mesmo tempo

durara	
Estándar ortográfico	durará
Lema estándar	durar
Clase de palabra (estándar)	Verbo (VMIF350) Main; indicative; future; third; singular
Tipo de desviación del estándar	O_ac_om

Con respecto ao obxecto da estandarización e da codificación, normalízanse e etiquétanse aquelas formas, usos ou omisións que non corresponden ao código normativo galego, que son inadecuados para un rexistro formal e académico como o que corresponde ao das probas ABAU ou que non respectan as convencións establecidas para textos deste tipo, atendendo aos niveis ortográfico, morfolóxico, semántico, léxico, gramatical (sintáctico) e discursivo. En §2.5.3 pode atoparse unha descrición máis detallada dos aspectos que son estandarizados e etiquetados no corpus.

2.5.3. Un sistema multinivel

Tal e como sinalan Stemle et al. (2019, p. 441), «Although very few learner corpus projects have managed to reuse each other's error taxonomies so far, several projects have tried to build on previous work». *CORTEGAL* identifícase perfectamente con esta afirmación, posto que, se ben o seu sistema de estandarización e codificación das formas non estándar toma como punto de partida sistemas empregados noutros corpus, utiliza un método propio, construído non só sobre a base dos sistemas anteriores, senón tamén a partir da análise dos problemas detectados nos textos de *CORTEGAL* e tendo en conta as especificidades e posibilidades de *TEITOK*.

Unha das principais diferenzas entre os distintos corpus ten que ver cos niveis cubertos polas taxonomías empregadas na anotación das formas desviantes, entre os cales o ortográfico, o gramatical e o léxico adoitan estar presentes:

the linguistic levels more commonly covered by the error taxonomies are spelling, grammar and lexis. On the other hand, classifications of phonetic, pragmatic or discursal errors do not seem to be always present in error tagging systems, and when present, their error categories are rather limited (Díaz-Negrillo e Fernández Domínguez, 2006, p. 89).

No caso particular de *CORTEGAL*, o número de niveis establecidos é de seis, como xa adiantamos e como comentaremos con máis detalle nesta epígrafe. Ademais, o sistema de estandarización utilizado en *CORTEGAL* é, como noutros corpus (por exemplo, *COPLE2*, Mendes et al., 2016; *KoKo*, Abel et al., 2016; ou *Falko*, Reznicek et al., 2013) un sistema multinivel, que permite a asignación de formas normalizadas en diferentes dimensións lingüísticas (Díez-Bedmar, 2021, p. 97).

Algúns corpus multinivel, como o *CzeSL-man* (Rosen, 2015), *Falko* (Reznicek, 2012) ou *MERLIN* (MERLIN project, 2014) permiten a asignación de *target hypotheses* tan só en dous niveis. En trazos xerais, no primeiro nivel lévanse a cabo estandarizacións de carácter esencialmente formal, como as ortográficas e as gramaticais cando afectan a elementos illados, mentres que no segundo nivel se realizan outras modificacións máis complexas que teñen en conta o contexto, entre elas as semánticas e as estilísticas.

Noutros corpus multicapa, como o *CroLTec* (Mikelić Preradović, 2020) ou o *COPLE2* (del Río e Mendes, 2018) escóllense os tres niveis «clásicos» indicados por Díaz-Negrillo e Fernández Domínguez na cita anterior: ortográfico, gramático e léxico. Outros engaden un cuarto nivel relativo aos aspectos discursivos, como o corpus *KoKo*, cunha capa para a estandarización da puntuación.

Outro aspecto que pode diverxer entre os distintos corpus ten que ver coa distribución dos tipos de desviación entre os distintos niveis: por citar un exemplo, o corpus *KoKo*, tal e como indicamos, asigna un nivel específico para os problemas de puntuación, mentres que outros corpus, por exemplo o *COPLE2*, os integra dentro da ortografía (Amaro et al., 2020, p. 14).

De acordo co indicado, non existe unha proposta única de niveis lingüísticos nos corpus multicapa. No caso de *CORTEGAL*, tal e como indicamos, decidimos establecer seis niveis distintos, ordenados do seguinte xeito: ortográfico, morfolóxico, léxico, semántico, gramatical (sintáctico) e discursivo. O elevado número de niveis lingüísticos establecidos xustifícase pola posibilidade de que unha mesma forma reciba formas normalizadas distintas e incompatibles entre si en varios deses niveis, tal e como amosaremos máis abaixo nesta mesma epígrafe.

No nivel ortográfico, anotamos problemas relativos ao emprego de signos diacríticos (acentuación e diérese), de maiúsculas e minúsculas, á escritura conxunta ou separada das palabras, á confusión, adición ou omisión de letras ou dígrafos, á escritura de estranxeirismos e abreviaturas ou siglas e á representación de contraccións e asimilacións. Sobre os problemas de acentuación, véxase o capítulo de López-Sández e Lorenzo-Herrera neste volume.

No que respecta ao nivel morfolóxico atendemos exclusivamente a problemas de flexión, sexa de verbos, de substantivos, de adxectivos ou de palabras gramaticais. Deste xeito, estandarízanse e codifícanse neste nivel cuestións relativas ao xénero, ao número, á conxugación verbal e aos morfemas avaliativos.

No nivel léxico, estandarizamos e etiquetamos os casos de emprego dunha unidade léxica con diferenzas de xénero ou de acentuación de intensidade con respecto ao galego estándar, en ambos os casos xeralmente por influencia do cognado español (por exemplo, *leite*, con 15 exemplos en feminino por influencia do español *leche*, que ten este xénero; ou *élite*, con 7 exemplos, por influencia do español *élite*, fronte ao galego estándar *elite*). Pero, sobre todo, inclúense aquí outras desviacións léxicas en que as diferenzas co estándar van máis alá de modificacións no xénero ou na acentuación, como o uso de *receta* en vez de *receita*, de *abandoar* no canto de *abandonar*, de *actitud* por *actitude*, de *todavía* por *aínda* etc. Sobre as formas non estándar do nivel léxico *vid.* o capítulo de Álvarez de la Granja nesta obra.

No nivel gramatical atendemos a cuestións relacionadas coa sintaxe da frase e da oración: problemas de concordancia, omisión e adición de preposicións, problemas no emprego de pronomes, determinantes e conxuncións, selección gramaticalmente inadecuada de tempo, modo ou voz verbal, construción non estándar dun esquema verbal, escolla dunha clase de palabra inadecuada ou, en xeral, presenza dunha estrutura sintáctica agramatical. Véxanse o capítulo de Fernández Salgado sobre a colocación dos pronomes átonos e o de Cidrás sobre cuestións de concordancia, neste volume.

No nivel semántico anotamos os usos ou significados dunha palabra que non son estándar ou adecuados ao contexto ou ao sentido que plausiblemente se quere transmitir (por exemplo, o emprego de *causas* por *consecuencias* en «provocar causas nefastas»), así como a omisión de palabras necesarias para o sentido do texto ou a adición de voces ou expresións innecesarias, normalmente pola súa redundancia.

Finalmente, no nivel discursivo rexistramos os problemas relativos á puntuación, incluíndo a división do texto en parágrafos, e ao emprego de conectores e de partículas referenciais, así como as elipses inadecuadas, a orde discursivamente inapropiada, aínda que gramaticalmente aceptable, os problemas relativos ao rexistro das palabras ou construcións e, finalmente, os enunciados demasiado complexos e os inintelixibles.

Con respecto aos problemas de puntuación, que nalgúns corpus, como *COPLE2* ou *MERLIN*, se consideran dentro do nivel ortográfico, optamos

por incluílos no nivel discursivo na medida en que os signos de puntuación son un elemento fundamental para a organización do texto. Tal e como sinala Figueras (1999), «cabe entender la puntuación como un mecanismo más de organización de la información del texto; su función es delimitar y articular las diversas unidades textuales de procesamiento». Do mesmo xeito, Roselló Verdeguer (2015), quen considera os signos de puntuación como elementos de cohesión discursiva, incide na necesidade de separar ortografía e puntuación:

Los signos de puntuación han recibido poca atención por parte de la lingüística, que los ha considerado un aspecto de la ortografía a caballo entre la oralidad y la escritura. Aunque es cierto que la primitiva función de los signos era señalar los lugares donde el lector debía realizar las pausas, hoy se considera la puntuación como un elemento de la escritura, y, por tanto, un mecanismo para articular el contenido del texto y dotarlo de una estructura.

Tal e como indicamos, integramos, pois, os problemas de puntuación cos que afectan a outros mecanismos de cohesión, que, canda a puntuación, contribúen a dotar o texto de coherencia, como a conexión, as elipses e a referencia. Somos conscientes, en calquera caso de que ás veces os problemas de puntuación responden a unha mera cuestión ortográfica (por exemplo, a omisión dun punto ao final do texto) ou de que outras veces os signos de puntuación teñen unha dimensión máis gramatical ca discursiva, ao contribuír, por exemplo, á organización da frase. En calquera caso, e para non complicar en exceso o proceso de anotación, todos os problemas relativos á puntuación (agás a omisión de punto en abreviaturas, que se anota no nivel ortográfico coa etiqueta O_ab_su), son codificados na capa discursiva.

Tras establecer o número de capas de estandarización que interesaba no deseño xeral do corpus, a normalización das formas levouse a cabo, na maior parte dos casos (*vid.* §2.5.4 para excepcións), nun formulario de edición de *tokens* mediante a asignación, no nivel correspondente, da forma estandarizada. Así, por exemplo, retomando o exemplo de *durara* que amosamos na Figura 12, e dado que nos atopamos cun problema de acentuación gráfica, debemos escribir a forma correcta na capa de estandarización ortográfica, tal e como se ve na Figura 13. Ademais de levar a cabo a corrección, asignamos un código que identifica o tipo de desviación (*vid.* §2.5.4). A forma estandarizada, e o código correspondente, convértense en atributos do *token*, tal e como indicamos en §2.2.

FIGURA 13. Normalización e codificación en *TEITOK*

Token value (w-169): durara

pform	Transcription (Inner XML)	durara
form	Student final version	
ocform	Orthographic standard	durará
mcform	Morphological standard	
lform	Lexical standard	
gform	Grammatical standard	
sform	Semantic standard	
dform	Discursive standard	
<hr/>		
lemma	Standard lemma	
olemma	Original lemma	
pos	POS tag (standard)	gramaticais
opos	POS tag (original)	gramaticais
problem	Type of deviation of the standard	O_ac_om
psource	Source of the non-standard form	
dcorrection	Derived correction	
arg	Connector	

Nótese que, se non se indica nada nas capas de estandarización, estas van herdar sempre a última forma normalizada e, de non haber normalización de ningún tipo, a forma escrita no orixinal. Deste modo, neste caso, os niveis morfolóxico, léxico, gramatical, semántico e discursivo herdan a forma corrixida ortograficamente *durará*.

Tal e como sinalamos, a utilización dun sistema multinivel permite a asignación de varias formas normalizadas en capas distintas. Nestes casos, sempre que é posible e pertinente, a estandarización realizada nun nivel mantense nos seguintes. Así, por exemplo, no fragmento «o caldo galego era o medio de subsistencia das clases mais desfavorecidas, votando os restos das comidas no puchero», substituímos no nivel ortográfico *votando* por *botando* e no nivel gramatical, en que se muda o xerundio por unha oración de relativo, conservamos a corrección ortográfica («que botaban» e non «que votaban»), tal e como se ve na Figura 14.

FIGURA 14. Dobre estandarización e conservación das modificacións en *TEITOK*

Token value (w-104): votando

pform	Transcription (Inner XML)	votando
form	Student final version	
ocform	Orthographic standard	botando
mcform	Morphological standard	
lcform	Lexical standard	
gcform	Grammatical standard	que botaban
scform	Semantic standard	
dcform	Discursive standard	
<hr/>		
lemma	Standard lemma	
olemma	Original lemma	
pos	POS tag (standard)	gramaticais
opos	POS tag (original)	gramaticais
problem	Type of deviation of the standard	O_cons_su,G_vmt_su
psource	Source of the non-standard form	
dcorrection	Derived correction	
arg	Connector	

Agora ben, nalgúns casos tal conservación non é posible na medida en que as diferentes formas normalizadas non son compatibles entre si. Así, por exemplo, tal e como se ve na Figura 15, *esten*, terceira persoa de plural do presente do subxuntivo do verbo *estar*, corríxese no nivel ortográfico mediante *estén*, dado que é unha palabra aguda rematada en <-n> que debería levar acento gráfico, pero estandarízase de novo, desta volta mediante *estean*, na capa morfolóxica, dado que *estén* é unha forma flexionada non estándar, producida por influencia do español. Asignamos ademais os códigos correspondentes que identifican o tipo de desviación e a orixe da diverxencia morfolóxica, tal e como se explicará en §2.5.4.

FIGURA 15. Dobre estandarización sen conservación das modificacións en *TEITOK*

Token value (w-51): <i>esten</i>		
pform	Transcription (Inner XML)	<input type="text" value="esten"/>
form	Student final version	<input type="text"/>
ocform	Orthographic standard	<input type="text" value="estén"/>
mcform	Morphological standard	<input type="text" value="estean"/>
lcform	Lexical standard	<input type="text"/>
gcform	Grammatical standard	<input type="text"/>
scform	Semantic standard	<input type="text"/>
dcform	Discursive standard	<input type="text"/>
<hr/>		
lemma	Standard lemma	<input type="text"/>
olemma	Original lemma	<input type="text"/>
pos	POS tag (standard)	<input type="text" value="gramaticais"/>
opos	POS tag (original)	<input type="text" value="gramaticais"/>
problem	Type of deviation of the standard	<input type="text" value="O_ac_om,M_v_su"/>
psource	Source of the non-standard form	<input type="text" value="M_sp"/>
dcorrection	Derived correction	<input type="text"/>
arg	Connector	<input type="text"/>

Nótese que a forma *estean* non mantén, por incompatible, a corrección ortográfica relativa á acentuación e que se traballásemos nun sistema cun só nivel de *target hypotheses*, estandarizando directamente *esten* por *estean*, sen ofrecer como paso intermedio a forma normalizada *estén*, estaríamos agochando un paso fundamental da análise, aquel que xustamente explica a asignación do código de omisión de acento (O_ac_om). Precisamente, e tal e como indicamos antes, a proposta de seis niveis lingüísticos xustifícase polo feito de que cada un deles pode requirir estandarizacións diferenciadas e incompatibles, o que nos leva a distinguir, por exemplo, entre o nivel morfolóxico (flexión) e o gramatical (sintaxe da frase e a oración), niveis que nalgúns corpus, como *KoKo* e *MERLIN*, se unifican. No exemplo que se ofrece a continuación, encontramos a acumulación nunha mesma forma de problemas dos dous tipos, aos que corresponden formas normalizadas diferentes en cada nivel:

- (4) Tamén hai que ter en conta que no noso caso, en España, dende fai uns poucos anos na televisión *esteñan* a desenrolar programas exclusivamente de concursos de cociña

En (4), no nivel morfolóxico débese substituír a variante dialectal *esteñan* do presente de subxuntivo pola forma estándar *estean*, mentres que no nivel gramatical existe un problema na selección do modo verbal, de modo que é necesario empregar o indicativo: *están*. A substitución directa por *están* agocharía a forma normalizada *estean* que xustifica a asignación da etiqueta morfolóxica M_v_su.

Polo mesmo motivo, e fronte á práctica relativamente habitual nos corpus de aprendentes (por exemplo *COPLE2* ou *MERLIN*), diferenciamos entre o nivel léxico e o semántico, posto que en varios casos é preciso atribuír formas normalizadas diferentes e incompatibles para cada nivel, tal e como ocorre en (5):

(5) Os costes de produción son moi *amplios*

No fragmento anterior, o castelanismo *amplios* debe estandarizarse no nivel léxico por *amplos*, pero no nivel semántico é preciso levar a cabo outra modificación, posto que *amplos/amplios* non é un adxectivo apropiado ao contexto (fronte a, por exemplo, *elevados*, que empregamos como forma normalizada).

Os problemas de rexistro, codificados nalgúns corpus no nivel léxico (por exemplo no corpus *KoKo*), son tamén separados deste por idénticas razóns. Así, por exemplo, en (6)

(6) Teñen unha pinta *asquerosa*

cómpre estandarizar o castelanismo *asquerosa* por *noxenta* no nivel léxico, pero consideramos necesaria unha segunda modificación no nivel discursivo, na medida en que nin *noxenta* nin *asquerosa* parecen formas adecuadas ao rexistro formal dos textos de *CORTEGAL*, polo cal propoñemos *desagradable*.

TEITOK, por outro lado, permite visualizar os textos nas diferentes capas de estandarización, tendo en conta que en cada capa se ven as modificacións propias dese nivel e as dos anteriores. Así pois, na última capa, a discursiva, obtemos o texto estandarizado en todas as dimensións lingüísticas. Ademais, as formas normalizadas de cada nivel teñen asignadas en *CORTEGAL* cores distintas para poder distinguir claramente as unhas das outras. Así, por exemplo, na Figura 16 visualizamos un dos textos tras premer na capa de estandarización morfolóxica («Estándar morfolóxico» en «Opcións de visualización»), de tal xeito que vemos as modificacións realizadas no nivel

morfolóxico (un único cambio, en laranxa) e no nivel anterior, o ortográfico (en lila):

FIGURA 16. Visualización dun texto na capa de estandarización morfolóxica

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** **Estándar ortográfico** **Estándar morfolóxico** **Estándar léxico**
Estándar gramatical **Estándar semántico** **Estándar discursivo**

Mostrar: **Cores** **Aliñación** **<pb>** **<lb>**

Anotación: **Lema estándar** **Lema orixinal** **Clase de palabra (estándar)** **Tipo de desviación do estándar**
Fonte da forma non estándar **Corrección derivada** **Conector**

Nesta sociedade a produción e consumismo chean as nosas vidas. Calquera persoa que produce e gana o seu beneficio a cambio o que vai facer é consumir.

Na sociedade na cal vivimos, é, sen dúbida, unha sociedade consumista; xa que dende que nacemos ata que morremos non paramos de consumir. O peor deste tema é que as persoas compren cousas por **satisfacción** persoal, e para elas as fai sentir mellor; e non compran por necesidade.

É normal que calquera persoa teña algún que outro capricho; pero isto vai máis alá dun capricho. Moita xente "compra por comprar" cousas que verdaderamente non necesita e falta ao **respecto** ás persoas que teñen problemas para comprar comida. (nec. primaria).

Está claro que para que o sistema da economía funcione, **ten** que haber xente que produza pero tamén consuma, xa que é todo unha cadea. Pero tamén hai que ter en conta onde se invirte o diñeiro, porque hai que estar informado e ter en conta canta xente non pode ter esa sorte; e non comprar cousas que logo as vas a deixar tiradas e non as vas a utilizar.

En cambio, na Figura 17, o mesmo texto visualízase na capa discursiva (tras premer en «Estándar discursivo»), de tal xeito que ademais das formas normalizadas anteriores, vemos as modificacións do nivel léxico (en verde), do gramatical (en salmón), do semántico (en azul) e do discursivo (en fucsia).

FIGURA 17. Visualización dun texto na capa de estandarización discursiva

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** **Estándar ortográfico** **Estándar morfolóxico** **Estándar léxico**
Estándar gramatical **Estándar semántico** **Estándar discursivo**

Mostrar: **Cores** **Aliñación** **<pb>** **<lb>**

Anotación: **Lema estándar** **Lema orixinal** **Clase de palabra (estándar)** **Tipo de desviación do estándar**
Fonte da forma non estándar **Corrección derivada** **Conector**

Nesta sociedade a produción e o consumismo **enchen** as nosas vidas. Calquera persoa que produce e **obtén** o seu beneficio a cambio o que vai facer é consumir.

A sociedade na cal vivimos é, sen dúbida, unha sociedade consumista, xa que dende que nacemos ata que morremos non paramos de consumir. O peor **desta situación** é que as persoas compren **produtos** por **satisfacción** persoal, e a elas **fainas** sentir mellor; e non compran por necesidade.

É normal que calquera persoa teña algún que outro capricho; pero isto vai máis alá dun capricho. Moita xente "compra por comprar" **produtos** que **verdadeiramente** non necesita e falta ao **respecto** ás persoas que teñen problemas para comprar comida (**necesidade** primaria).

Está claro que para que o sistema da economía funcione, **ten** que haber xente que produza pero tamén consuma, xa que é todo unha cadea. Pero tamén hai que ter en conta onde se **inviste** o diñeiro, porque hai que estar informado e ter en conta canta xente non pode ter esa sorte; e non comprar **produtos** que logo vas deixar **tirados** e non vas utilizar.

Do exposto ata o de agora dedúcese que a orde establecida para as capas de estandarización non é en absoluto irrelevante, posto que as modificacións se realizan e se visualizan de acordo con esa orde, comezando polas mudanzas formais a nivel de palabra (ortografía, morfoloxía e léxico: primeiro as meramente gráficas, despois as que afectan á flexión e finalmente as que corresponden á selección da unidade léxica), seguindo polas formais a nivel da frase ou da oración (gramática), continuando coas relacionadas co significado (semántica) e rematando con aquelas que están vinculadas coa construción do texto e coa adecuación ao contexto (discurso).

2.5.4. *Os códigos de anotación das formas non estándar*

Tal e como xa comentamos e como queda patente en numerosos traballos (véxase, por exemplo, Díaz-Negrillo e Fernández Domínguez, 2006; Stemle et al., 2019 ou Díaz-Bedmar, 2021), a taxonomía para a clasificación das formas desviantes do estándar que atopamos naqueles corpus de aprendentes que ofrecen este tipo de anotación pode variar considerablemente duns proxectos a outros. Estas diferenzas encóntranse nos criterios de clasificación empregados, na granularidade da taxonomía ou nos niveis ou aspectos lingüísticos considerados.

As clasificacións e estudos realizados en diferentes traballos antes do nacemento da análise informatizada de erros (entre eles Corder, 1974; Dulay et al., 1982 ou Ellis, 1994) serviron como base para as taxonomías aplicadas nos corpus informatizados. Deste xeito, é frecuente atopar clasificadas as formas non estándar atendendo ao nivel lingüístico ou ao aspecto concreto ao que afecta a desviación e tamén tendo en conta as modificacións que se producen con respecto ao estándar (omisións, adicións, cambios de orde...). Menos frecuentemente, tamén se poden encontrar clasificacións que consideran as causas da desviación (erros interlingüísticos, intralingüísticos...).

No caso de *CORTEGAL*, a clasificación das formas desviantes do estándar realízase atendendo aos tres aspectos mencionados, en dúas taxonomías separadas. A primeira tenta describir a forma non estándar combinando información sobre o nivel e o aspecto afectado con información sobre as modificacións que se producen con respecto ao estándar e a segunda, só empregada en certos casos, dá conta da orixe ou causa da desviación.

Na primeira das taxonomías, empregamos códigos conformados por tres elementos separados por barra baixa, como, por exemplo, O_ac_om. O primeiro elemento, sempre en maiúscula, indica o nivel lingüístico ao que afecta a desviación: O: Ortografía, M: Morfoloxía, L: Léxico, G: Gramática (sintaxe), S:

Semántica e D: Discurso. O segundo elemento, o que ofrece maior variación, concreta o aspecto afectado dentro do nivel lingüístico xeral (*ac*: acentuación, *gen*⁷: xénero, *ref*: elementos con valor referencial, *prep*: preposicións etc.). Como se pode observar, tales elementos poden ser clases de palabras, trazos lingüísticos, categorías funcionais etc. Cando queremos indicar unha unidade léxica en xeral usamos «w» («word»), como en L_w_su (substitución dunha unidade léxica estándar por unha non estándar). Finalmente, o terceiro elemento explica a diferenza existente entre a forma estándar e a forma proporcionada pola/o estudante, sempre desde esta última perspectiva, ou o trazo que fai que a forma ou secuencia escrita non sexa estándar (*ad*: adición, *om*: omisión, *su*: substitución, *wp*: posición incorrecta ou inadecuada, *tr*: transposición, *un*: unión, *spl*: división, *com*: excesiva complexidade, *unin*: inintelixibilidade).

O número total de códigos é de 80, distribuídos do xeito que se indica na Táboa 3, onde se pode observar que os niveis ortográfico, gramatical e discursivo son os que contan cun maior número de etiquetas. A listaxe completa de códigos, con explicacións, exemplos e aclaracións sobre o seu uso, pode atoparse no manual empregado para a etiquetaxe e estandarización e que se encontra dispoñible na páxina do proxecto (<http://ilg.usc.gal/cortegal/downloads/manual-anotacion.pdf>).

TÁBOA 3. Distribución dos códigos descritivos das formas non estándar por nivel lingüístico

Nivel	Número de códigos	Porcentaxe
Ortográfico	25	31,25%
Gramatical	24	30%
Discursivo	18	22,5%
Morfolóxico	6	7,5%
Semántico	4	5%
Léxico	3	3,75%

Somos conscientes de que se trata dun número elevado de códigos e de que este feito dificulta a anotación e pode afectar á súa coherencia (*vid.* Díaz-Bedmar, 2021, p. 92). Con todo, tales códigos están construídos mediante a combinación dun número máis reducido de elementos (6 do primeiro nivel, 37 do segundo e 9 do terceiro), o que, no sentido contrario, facilita a anotación.

⁷ As abreviacións están creadas sobre as formas inglesas, pensando na posible replicabilidade dos códigos en corpus doutras linguas.

Ademais, dado que o corpus está pensado non só para a investigación, senón tamén para o seu emprego directo nas aulas, unha etiquetaxe demasiado xeral podería ser pouco útil neste contexto. Optamos, daquela, por unha maior granularidade que permita unha clasificación máis desenvolvida das formas desviantes e, daquela, que facilite a localización por parte do profesorado e do alumnado de exemplos de desviacións relativamente concretas.

Como xa indicamos en §2.2, a maior parte das anotacións que estamos comentando son realizadas a nivel de *token*, mediante o mesmo formulario empregado para a anotación das formas normalizadas que presentamos en §2.5.3. Unha mesma forma que acumule varias desviacións recibirá varios códigos, pertencentes ao mesmo ou a distintos niveis, tal e como se ve na Figura 18, en que, na palabra *Hostelería*, no fragmento «a Hostelería gústalle a xente» se acumulan dúas anotacións no nivel ortográfico (omisión de acento gráfico [O_ac_om] e adición de maiúscula [O_uc_ad]) e unha no nivel léxico (emprego de *hostelería* por *hostalería*) [L_w_su].

FIGURA 18. Asignación de varios códigos de forma non estándar a unha mesma palabra

Token value (w-164): Hosteleria

pform	Transcription (Inner XML)	Hosteleria
form	Student final version	
ocform	Orthographic standard	hosteleria
mcform	Morphological standard	
lcform	Lexical standard	hostalería
gcform	Grammatical standard	
scform	Semantic standard	
dcform	Discursive standard	
lemma	Standard lemma	
olemma	Original lemma	
pos	POS tag (standard)	gramaticais
opos	POS tag (original)	gramaticais
problem	Type of deviation of the standard	O_ac_om,O_uc_ad,L_w_su
psource	Source of the non-standard form	L_sp
dcorrection	Derived correction	
arg	Connector	

Nótese, porén, que esta multiplicidade de códigos está asociada sempre á acumulación de diferentes desviacións, posto que para cada unha delas

ofrecemos un único código e nunca varias interpretacións (sobre esta cuestión, *vid.* §2.5.6).

Un total de 7 dos 80 códigos arriba mencionados non se anotan a nivel de *token*, senón en arquivos separados ligados ao texto, empregando un sistema *stand-off*. Isto sucede con aqueles códigos que afectan ou poden afectar a unha secuencia de *tokens*, que son os que identifican as reformulacións e topicalizacións que dan lugar a estruturas gramaticais inaceptables (G_str_ref-top_su), as adicións de palabras que orixinan tamén unha estrutura agramatical (G_w_ad), as estruturas agramaticais que non encaixan noutros códigos (G_str_su), o cambio de orde discursivamente inadecuado (D_w_wp), os enunciados excesivamente complexos (D_ut_com) e os inintelixibles (D_ut_unin), así como a adición de elementos innecesarios, normalmente por seren semanticamente redundantes (S_w_ad). A razón desta diferente anotación explícase na páxina de *TEITOK* e radica nas dificultades de asignar un código a toda a secuencia se os seus compoñentes xa posúen outras anotacións:

Whenever more information needs to be encoded in the corpus that spans multiple words, and which can with [overlap] other types of information, there is not way to encode such information inside the XML file. The only way to encode such information is what is called stand-off: stored in a separate file which is linked to the TEI document⁸.

Debe terse en conta que os códigos indicados se anotan sempre mediante un sistema *stand-off*, independentemente de que nalgúns casos poidan afectar unicamente a unha palabra. Así, por exemplo, a etiqueta S_w_ad pode atinxir a unha secuencia de palabras, como en (7), en que sobra o elemento parentético que figura ao final, pois resulta redundante, ou a unha única palabra, como en (8), onde sobra *quizá* (os destacados son nosos):

- (7) A gastronomía e a cociña, nos últimos anos acadaron moita popularidade, aínda que hai que recoñecer, que polo menos en Galicia sempre foi e continúa a ser dos trazos que máis caracterizan o noso patrimonio e cultura (*no ámbito gastronómico*)
- (8) É probable que *quizá* a estética dos platos evolucione

Só desta maneira aseguramos que o conxunto de exemplos codificado coa mesma etiqueta (neste caso con S_w_ad) se recupere conxuntamente, posto que os códigos asignados mediante este sistema de anotación deben ser buscados nun apartado específico do buscador de *TEITOK* («Anotacións multipalabra»).

⁸ <http://www.teitok.org/index.php?action=help&id=standoff>

Por outro lado, tal e como se ve na Figura 20, as estandarizacións realizadas mediante un sistema *stand-off* son visibles premendo na ligazón «Anotación multipalabra» que aparece naqueles textos que contan con este tipo de anotacións (destacado cun rectángulo vermello na Figura 19):

FIGURA 19. Ligazón para consultar as anotacións multipalabra

Opcións de visualización

Texto: **Transcripción completa** | **Versión final estudante** | Estándar ortográfico | Estándar léxico | Estándar gramatical

Estándar semántico | Estándar discursivo

Mostrar: **Cores** | **Alliñación** | <pb> | <lb>

Anotación: Lema estándar | Lema orixinal | Clase de palabra (estándar) | Clase de palabra (orixinal)

Tipo de desviación do estándar | Fonte da forma non estándar | Corrección derivada | Conector

A gastronomía actualmente é unha profesión enormemente coñecida por todas as partes do mundo, onde non todos podemos destacar, xa que realizar unha comida con presentación e unha sabor perfectas non é sinxelo.

Podemos coñecer moitos actores, e o cine xa é unha profesión onde case ninguén pode destacar, pero aínda menos coñecemos a gastrónomos famosos, que faría falla unha vida de práctica para alcanzar a cociña sen erros.

A súa gran importancia relacionase cos países e coa cultura, xa que cada país ten a súa propia gastronomía e pratos especiais, coma a súa danza ou o seu idioma.

É coma unha das pequenas partes dun país e da súa cultura natal. Aquí moitas persoas obteñen curiosidade ou aprecio pola gastronomía de outros lugares. Probar a comida dun país nun viaxe é tan común coma observar o idioma que ali se fala, polo tanto penso que o aprecio a este ámbito cultural non se vai perder cos anos, pode incluso millorar. É probable que quizá a estética dos platos evolucione, pero o contido dos platos sempre representará o mesmo tras o [---] paso do tempo.

A gastronomía realmente é fermosa e complicada, unha profesión onde non calquera pode destacar e facer **só un** esforzo mínimo. E algo digno de admirar; tanto en Galicia coma na outra punta do mundo.

Lenda: **Lectura difícil** • Texto borrado • **Texto engadido**

Descargar vista actual como txt: **Anotación multipalabra**

FIGURA 20. Pantalla que se visualiza tras premer na ligazón Anotación multipalabra

Anotación multipalabra

Annotation of multi-token units.

A gastronomía actualmente é unha profesión enormemente coñecida por todas as partes do mundo, onde non todos podemos destacar, xa que realizar unha comida con presentación e unha sabor perfectas e non é sinxelo.

Podemos coñecer moitos actores, e o cine xa é unha profesión onde case ninguén pode destacar, pero aínda menos coñecemos a gastrónomos famosos, que faría falla unha vida de práctica para alcanzar a cociña sen erros.

A súa gran importancia relacionase cos países e coa cultura, xa que cada país ten a súa propia gastronomía e pratos especiais, coma a súa danza ou o seu idioma.

É coma unha das pequenas partes dun país e da súa cultura natal. Aquí moitas persoas obteñen curiosidade ou aprecio pola gastronomía de outros lugares. Probar a comida dun país nun viaxe é tan común coma observar o idioma que ali se fala, polo tanto penso que o aprecio a este ámbito cultural non se vai perder cos anos, pode incluso mellorar: **É probable que quizá** a estética dos platos evolucione, pero o contido dos platos sempre representará o mesmo tras o paso do tempo.

A gastronomía realmente é fermosa e complicada, unha profesión onde non calquera pode destacar e facer só un esforzo mínimo. E algo digno de admirar, tanto en Galicia coma na outra punta do mundo.

Anotacións

Linguistic area

Semantics

• É probable que quizá

É probable que quizá	
Área lingüística	Semántica (Semantics)
Código	S_w_ad
Forma estándar	É probable que

Porén, estas estandarizacións non se actualizan nas distintas capas de visualización mencionadas en §2.5.3. Por tal motivo, e co obxectivo de que estas capas ofrezan unha versión estándar completa do texto, ademais de informar da forma modificada en *stand-off*, levamos a cabo a nivel de *token* a mudanza individual de cada unidade afectada (sen asignar códigos, evidentemente, que corresponden a toda a secuencia e só son asignables en *stand-off*), ata obter a versión final normalizada. Por este motivo, algunhas palabras que aparecen estandarizadas nos textos non están asociadas, a nivel de *token*, a ningún tipo de código.

A anotación en *stand-off*, limitada aos códigos indicados, debe diferenciarse da anotación de unidades complexas creadas en *TEITOK* mediante a función *merge* que ofrece a plataforma, e que empregamos cando é preciso estandarizar e codificar en bloque unha expresión complexa (pode ser unha expresión lexicalizada ou calquera outra unidade que realice unha función unitariamente e deba ser substituída en conxunto). Así, por exemplo, unificamos os compoñentes da locución substantiva *perrito quente*, que se estandariza e etiqueta conxuntamente con *bocadillo de salchicha* e *L_w_su*. En calquera caso, restrinximos o máximo posible o emprego desta función, que ofrece algunhas limitacións e certas dificultades na recuperación a través do buscador.

A anotación que comentamos ata o de agora serve para describir o tipo de desviación con respecto ao estándar, pero, tal e como indicamos ao comezo deste apartado, existe unha segunda taxonomía, moito máis simple, que ten como función ofrecer posibles explicacións sobre a diverxencia da norma. Porén, este tipo de anotación é moito máis subxectiva, tal e como sinalan Lüdeling et al. (2005), e en moitas ocasións resulta realmente difícil determinar a orixe da forma desviante. Por tal motivo, a aplicación deste código segue criterios diferentes dos empregados no uso do código descritivo. Por un lado, non todas as formas desviantes reciben unha etiqueta explicativa e, por outro lado, en ocasións ofrecemos máis dun código deste tipo para explicar unha mesma desviación. As etiquetas empregadas, os niveis en que se aplica, o seu valor e algúns exemplos figuran na Táboa 4:

TABOIA 4. Códigos empregados para explicar a orixe da forma diverxente

Orixe	Valor	Códigos	Exemplos
<i>anal</i>	Formas que responden a un proceso de analoxía con outras palabras da mesma lingua	O_anal M_anal L_anal	O_anal: <i>hippie</i> (sobre o plural <i>hippies</i>)... M_anal: <i>puidendo</i> (sobre o tema do pretérito <i>puid-</i>)... L_anal: <i>nutrinte</i> (a partir de voces como <i>ferinte</i> ou <i>seguinte</i>)...
<i>for</i>	Palabras cuxa desviación se pode explicar pola súa orixe estranxeira e calcos de voces estranxeiras. Exclúense sempre as formas con orixe no español. No ámbito ortográfico, palabras cunha ortografía incorrecta. No ámbito morfolóxico, palabras con flexión non estándar. No ámbito léxico, palabras que non presentan a adaptación proposta no estándar. No ámbito semántico, calcos semánticos.	O_for M_for L_for S_for	O_for: <i>graffitti</i> (por <i>graffiti</i>)... M_for: <i>chef</i> (por <i>chefs</i>) L_for: <i>tablet</i> (por <i>tableta</i>)... S_for: <i>pretender</i> (co significado de 'finxir')...
<i>gal</i>	Palabras, trazos ou usos transferidos desde unha variedade do galego non estándar. No ámbito morfolóxico, flexión propia do galego popular. No ámbito léxico, palabra tirada dunha variedade do galego (popular ou culto). No ámbito gramatical, uso propio dalgunha variedade dialectal. No ámbito semántico, uso tirado dunha variedade do galego (galego popular ou culto)	M_gal L_gal G_gal S_gal	M_gal: <i>consegueu</i> (por <i>conseguiu</i>)... L_gal: <i>atrais</i> (por <i>atrás</i>)... G_gal: uso de <i>lle</i> por <i>lles</i> , cheísmo e teísmo S_gal: <i>mirar</i> (co valor de <i>ver</i>)...
<i>hc</i>	Hipercorreccións xeradas dentro do galego	L_hc G_hc	L_hc: hiperenxebrismos como <i>vaciña</i> (por <i>vaciña</i>), <i>calqueira</i> ... G_hc: colocación enclítica do pronome en vez de proclítica e uso de <i>che</i> por <i>te</i>
<i>mix</i>	Mesturas entre dúas palabras ou expresións	L_mix	L_mix: <i>indixesta</i> (<i>indigestión</i> + <i>inxesta</i>), <i>a saber Deus</i> (<i>a saber</i> + <i>sabe Deus</i>), <i>adedicar</i> (<i>adicar</i> + <i>dedicar</i>)...

TÁBOA 4. Continuación

<i>or</i>	Unicamente se emprega no ámbito gramatical para marcar os casos de pronomes enclíticos posibles na lingua oral, pero non recomendados na escrita formal	G_or	G_or: colocación enclítica do pronome posible na lingua oral, pero non recomendada na escrita formal
<i>sp</i>	Palabra, expresión, trazo ou uso transferidos desde o español. No ámbito ortográfico, grafía coincidente co español ou acentuación gráfica explicable por influencia desta lingua. No ámbito morfolóxico, flexión coincidente co español. No ámbito gramatical, colocación proclítica do pronome, así como ausencia ou adición de determinante, adición de preposición ou de pronome e substitución de determinante ou de pronome explicables por influencia do español. No ámbito léxico, unidades léxicas, acentuación de intensidade ou xénero transferidos desde o español. No ámbito semántico, calcos semánticos do español	O_sp M_sp G_sp L_sp S_sp	O_sp: <i>ahí</i> (por <i>aí</i>), <i>él...</i> M_sp: <i>anglosaxones...</i> G_sp: <i>levarse o diñeiro</i> (por <i>levar o diñeiro</i>)... L_sp: <i>actitud, élite, (a) leite</i> S_sp: <i>pobo</i> (por <i>vila</i> ou <i>aldea</i>)...
<i>sp_adapt</i>	No ámbito morfolóxico, flexión adaptada desde o español. No ámbito léxico, palabras ou expresións transferidas desde o español e adaptadas ao galego, ou mesturas entre elementos do galego e do español	M_spadapt L_spadapt	M_spadapt: <i>atraxo, supoñe...</i> L_spadapt: <i>semexante, ventaxes, chear...</i>

Nótese que, máis alá daquelas formas diverxentes do estándar cuxa orixe se nos escapa, existe un uso restrinxido desta anotación. Isto é, non todas as formas para as que é posible ofrecer unha hipótese plausible relativa á súa orixe reciben un código deste tipo, senón só aquelas que se corresponden co exposto na Táboa 4: por exemplo, a marca de oralidade podería utilizarse, pero non se emprega, para etiquetar o uso da contracción *ca* en vez de *coa*. Son razóns de carácter práctico, derivadas da enorme complexidade do proceso de anotación, así como da procura da maior coherencia posible, as que explican este proceder. Por outro lado, debe terse en conta que o emprego das etiquetas é en certos casos bastante cauto. Así, por exemplo, en relación *coa* influencia do español na acentuación gráfica, só identificamos con O_sp os casos de adición e non os de omisión: a modo de ilustración, márcanse con este código os adverbios en *-mente* acentuados graficamente, pero non

as frecuentes omisións de acento na secuencia [u'i] (por exemplo *construir*, *incluír*, *ruido*), que tamén se poderían explicar polas diferenzas entre o sistema de acentuación do galego e do español (*vid.* o capítulo de López-Sández e Lorenzo-Herrera neste volume).

Lémbrese, por último, que, tal e como indicamos, un mesmo *token* pode recibir varios códigos explicativos. Así, por exemplo, os empregos cheístas van etiquetados tanto con G_hc como con G_gal, pois poden responder tanto a unha ultracorrección como a un uso dialectal. A introdución deste tipo de código realízase nun campo específico do mesmo formulario utilizado para a edición dos *tokens* presentado previamente (na Figura 18 pode encontrarse un exemplo).

2.5.5. As correccións derivadas

En *CORTEGAL* podemos encontrar varios exemplos de formas normalizadas que non están asociadas a un código identificador de forma desviante, aínda que si ao código DC («derived correction»). Son dous os tipos de circunstancias en que utilizamos esta anotación.

En primeiro lugar, asignamos o código DC cando é necesario realizar un cambio no texto como consecuencia doutra estandarización levada a cabo neste, dun modo parcialmente similar ao que se propón para o corpus *ASK*. Neste corpus utilízase unha etiqueta específica para os chamados «agreement errors»: «errors following logically from, and triggered by, previous errors, the agreement itself being in accordance with the target language norm» (Tenfjord et al., 2006, p. 1822). Existen con todo algunhas diferenzas importantes entre o sistema empregado no corpus *ASK* e en *CORTEGAL*: no corpus *ASK* os «agreement errors» considéranse como unha subcategoría de erro ligada ás categorías «Puntuación», «Maiúsculas e minúsculas» e «Selección errónea de categoría morfosintáctica», mentres que en *CORTEGAL* a marca DC non se considera un código de «forma non estándar» e ademais esténdese a calquera caso en que sexa necesario realizar unha estandarización derivada doutra.

Así, por exemplo, ao estandarizar o calco semántico do español *pobo* por *vila* na frase «o pobo de Iago Aspas», é necesario substituír o artigo masculino polo feminino para evitar un problema de concordancia («a vila de Iago Aspas»). Agora ben, o/a estudante non comete ningún erro de concordancia, de tal modo que carecería de sentido asignarlle a o un código de desviación gramatical. Nese caso asignamos no determinante o código DC e a forma normalizada «a», sempre no mesmo nivel no que se realiza a estandarización que a orixina (neste caso no nivel semántico), tal e como se pode comprobar na Figura 21.

FIGURA 21. Anotación dunha corrección derivada

Token value (w-184): o		
pform	Transcription (Inner XML)	<input type="text" value="o"/>
form	Student final version	<input type="text"/>
ocform	Orthographic standard	<input type="text"/>
mcform	Morphological standard	<input type="text"/>
lcform	Lexical standard	<input type="text"/>
gcform	Grammatical standard	<input type="text"/>
scform	Semantic standard	<input type="text" value="a"/>
dcform	Discursive standard	<input type="text"/>
<hr/>		
lemma	Standard lemma	<input type="text"/>
olemma	Original lemma	<input type="text"/>
pos	POS tag (standard)	<input type="text" value="gramaticais"/>
opos	POS tag (original)	<input type="text" value="gramaticais"/>
problem	Type of deviation of the standard	<input type="text"/>
psource	Source of the non-standard form	<input type="text"/>
dcorrection	Derived correction	<input type="text" value="DC"/>
arg	Connector	<input type="text"/>

Deste, xeito, a estandarización de *pobo* por *vila* vai acompañada da modificación de *o* por *a* e ambos os cambios son visualizados simultaneamente na capa de estandarización semántica, tal e como se ve na Figura 22, onde está activada esta última. O determinante modificado aparece en azul, como *vila*, pero non leva asignado ningún código identificador dun tipo de desviación, senón simplemente a marca DC.

FIGURA 22. Visualización das formas con corrección derivada

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico Estándar morfolóxico Estándar léxico
 Estándar gramatical Estándar semántico **Estándar discursivo**

Mostrar: **Cores** Aliñación <pb> <lb>

Anotación: Lema estándar Lema orixinal Clase de palabra (estándar) Clase de palabra (orixinal)
 Tipo de desviación do estándar Ponte da forma non estándar Corrección derivada Conector

A día de hoxe, é moi difícil encontrar un rapaz ou rapaza que, á hora de preguntarlle pola súa profesión cando sexa maior; non cambie cada certo tempo de resposta. Isto é comprensible, xa que co paso dos anos vas dándote conta de moitas máis cousas e descubrido novos oficios, chegando ata tal punto que rematas os teus estudos e aínda non sabes moi ben onde vas a parar. Pero o que non acabo de entender é a obsesión que mostran os rapaces de agora por chegar a ser futbolistas e as rapazas por chegar a ser modelo; aínda que isto ten unha explicación sinxela: tanto rapaces como rapazas a día de hoxe están dende pequenos involucrados coas novas tecnoloxías, sexa ordenador, televisión, móbil... Os rapaces queren ser o que se lles mostra nos medios de comunicación, que son futbolistas e modelos, pero pregúntome: onde están os bombeiros, médicos...? Estes non aparecen, e menos galegos, xa que se algo se emite sobre a cultura tratará sobre a vila de Iago Aspas ou Lucas Pérez. Antes os nenos querían ter o oficio da casa, o que lles amosaban os pais; agora xa non hai respecto por iso e tódolos rapaces se p futbolistas; en ningunha televisión se ve o oficio dos ferreiros, canteiros..., o que vai desaparecer. Estou en total acordo coa autora, xa que expón un problema que a maioría dos pais están acostumados a que soñen os seus fillos con ser futbolistas e modelos e estas nenas se dean conta da realidade.

Estes rapaces son o futuro da cultura galega e se non coñecen as orixes da súa cultura rapaces e rapazas teñan fillos? A cultura galega quedará no esquecemento.

Lenda: **Lectura difícil** • Texto borrado • Texto engadido

	o
Estándar semántico	a
Lema estándar	o
Clase de palabra (estándar)	Determinante (DAOMS0) Article; masculine; singular
Corrección derivada	DC

Un exemplo do mesmo tipo atopámolo naqueles casos en que propomos a inclusión dun punto, o que conduce á necesidade de empregar maiúscula inicial na palabra seguinte. Así, por exemplo, no seguinte fragmento,

- (9) Anteriormente, o turismo non estaba tan potenciado, habitualmente podemos observar numerosos anuncios por medios nos que se ofertan billetes de avión, barco etc. a bós prezos para que a xente se anime a viaxar e coñecer novos lugares

propoñemos substituír a coma que figura antes de *habitualmente* por un punto, o que obriga a utilizar maiúscula na primeira letra desta palabra, que pasa a etiquetarse con DC. Tamén se emprega o código DC, por exemplo, para a normalización das formas concordadas con palabras que reciben a etiqueta L_gen_su, a cal serve para indicar que se produce unha asignación de xénero non estándar a un substantivo. Sendo así, en (10), onde destacamos en cursiva as voces concordadas con *leite*,

- (10) A leite francesa resulta mais barata (o cal quere dicir que costa uns poucos céntimos menos) ca galega

non parece moi apropiado asignar códigos de desviación gramatical a tales voces, que concordan adecuadamente co xénero que a persoa autora do texto lle atribúe a *leite*. Daquela, todas esas formas son normalizadas no nivel

léxico (o mesmo en que se etiqueta *leite*) e anotadas con DC, pero non reciben ningún outro código que identifique unha desviación e que se compute como tal (sobre o código L_gen_su, e a súa diferenciación con G_gen_su, que marca un problema de concordancia, *vid.* §2.5.6 e o capítulo de Álvarez de la Granja neste volume).

Outros moitos casos implican a anotación con DC: a estandarización dos pronomes proclíticos que deberían colocarse como enclíticos (o código de desviación gramatical asígnaselle ao pronome mal colocado, que se suprime, mentres que ao verbo se lle engade o pronome enclítico, acompañando esta modificación de DC); o engadido dunha preposición ao levar a cabo unha estandarización no nivel léxico ou semántico dun verbo que non rexe preposición por outro que si o fai; a inclusión dunha forma léxica no canto dun clítico proclítico cun referente irrecuperable no texto (o clítico anótase con código de desviación e suprímese, mentres que a forma léxica se etiqueta con DC) etc.

O segundo tipo de circunstancias en que se emprega DC é aquel en que un mesmo e único problema ten varias manifestacións no texto. En tales casos, asignamos código de forma non estándar só nunha desas manifestacións e DC nas restantes. Un exemplo deste tipo atopámolo nos problemas de concordancia, que moitas veces se materializan en varias palabras. Así sucede, por exemplo, nos seguintes fragmentos:

- (11) Aínda que haxa xente que renege da cociña moderna porque *consideren* que son pratos cunha porción pequena ou que *estean acostumbrados* a cociña tradicional, hai que admitir o cambio da cociña e a gastronomía nos últimos anos
- (12) O capitalismo e o afán de poder *desaparecería* polo feito de non necesitalo

Nas tres palabras destacadas en cursiva en (11) hai concordancia *ad sensum*, e as tres voces deberían ser estandarizadas, respectivamente, por *considere*, *estea* e *acostumada* para lograr a concordancia gramatical con *xente*. Un problema do mesmo tipo figura en (12), onde o verbo e o pronome destacados deberían estar en plural pola súa concordancia cunha construción coordinada. Á hora da anotación temos dúas opcións: colocar unha etiqueta G_num_su en cada palabra, de tal modo que no caso de (11) se computarían tres problemas diferentes de concordancia de número e no caso de (12) dous, ou ben colocar a etiqueta na primeira delas e simplemente modificar as demais, pero sen asignarlles código de forma non estándar (en *acostumada*, con todo, si se incluíría un código de desviación relativo á concordancia de xénero).

Aínda que nestes exemplos sería xustificable a anotación de todas as palabras, optamos por etiquetar como desviante só unha delas. A razón estriba fundamentalmente no cómputo de formas non estándar: ao noso entender, en casos como (11) e (12) é máis adecuado considerar que existe un único problema de concordancia (un problema na concordancia con *xente* e con *o capitalismo e o afán de poder*, respectivamente), que ten diferentes manifestacións no texto, antes que varios problemas computables de maneira independente. Dado que, en calquera caso, todas as formas con problemas de concordancia son estandarizadas, ao consultar o texto, todas elas, tanto as anotadas con código de desviación como as que non o están, destacarán convenientemente coa cor correspondente (neste caso salmón, ao tratarse dunha estandarización realizada no nivel gramatical). Ademais, os *tokens* modificados pero non etiquetados con código de forma non estándar irán acompañados da etiqueta DC, que dá conta de que responden á segunda, terceira... manifestación dun problema xa anotado previamente, tal e como se pode comprobar na Figura 23.

FIGURA 23. Visualización dunha corrección derivada nun caso de concordancia

Opcións de visualización

Texto: **Transcripción completa** **Versión final estudante** Estándar ortográfico **Estándar morfolóxico** Estándar léxico

Estándar gramatical Estándar semántico Estándar discursivo

Mostrar: **Cores** **Aliñación** <pb> <lb>

Anotación: Lema estándar Lema orixinal **Clase de palabra (estándar)** Tipo de desviación do estándar

Fonte da forma non estándar Corrección derivada Conector

A gastronomía e a cociña alcanzaron o seu momento de auxe no século XXI grazas, en parte, á **súa** modernización. Aínda que haxa xente que **renegue** da cociña moderna porque **considere** que son pratos cunha porción pequena ou que **estea** **acostumada** á cociña tradicional, hai que admitir o cambio da cociña e a gastronomía nos últimos anos.

Pasou de ser un acto **cotián** **practicado** na casa a ser de programas de televisión como "MasterChef", no facer o mellor prato en cada fase ata **levar** un premio.

A nivel educativo tamén avanzou, xa que agora é p^{er} superiores, que **substitúen** o **bacharelato** e a **univer**...

As redes sociais e os medios **axudárona** como **axer** televisión, onde xa **conta** con numerosa audiencia, **Finalmente**, a cociña e a gastronomía **modernas** **dé** que os adornos e a parte artística son realmente d...

	estean	diais e incluso protagonista
Versión final estudante	estean	nha competición e tratan de mo un curso de cociña.
Estándar gramatical	estea	vos como os graos medios e
Lema estándar	estar	
Clase de palabra (estándar)	Verbo (VMSP3P0) Main; subjunctive; present; third; plural	, por exemplo , as cadeas de
Corrección derivada	DC	ue hai pratos elaborados nos

Por razóns do mesmo tipo, utilizamos o procedemento indicado no caso dos incisos, conectores... que deberían estar demarcados por dúas comas no texto, pero en que ambas faltan, como sucede coa oración de xerundio que figura en (13):

- (13) Grupos de homes, mulleres e nenos son obrigados construír empregando materiais tóxicos eses obxectos

Unicamente marcamos con código identificador de omisión de signo de puntuación (D_pm_om) a primeira das comas ausentes e incorporadas, mentres que a segunda coma se introduce exclusivamente coa marca DC. Consideramos que estamos ante un único problema (a ausencia de delimitación dun inciso) ao que lle corresponde unha dobre manifestación no texto e entendemos, en consecuencia, que debería computar unha soa vez. O mesmo sistema é empregado cando nun texto se deben incluír parénteses ou comiñas de apertura e peche.

2.5.6. Sobre a subxectividade da estandarización e da codificación en CORTEGAL

Como se sinala en moitos traballos (*vid.* por exemplo, Granger, 2003, p. 473; Díaz-Negrillo e Fernández Domínguez, 2006, p. 89; Reznicek et al., 2013, p. 104 ou Granger, 2017) a estandarización e a codificación das formas non estándar é un proceso complexo que encerra unha carga alta de subxectividade: «The subjectivity involved in identifying, correcting and classifying errors, which makes the process a time-consuming, effort-driven, and challenging task (...), is a characteristic of the error-tagging process» (Díaz-Bedmar, 2021, p. 95). A subxectividade na anotación de *CORTEGAL* faise presente de diferentes maneiras e con implicacións de maior ou menor relevancia.

Por un lado, en ocasións existen varias posibilidades para escoller *target hypothesis*, como no caso de *sin embargo* ou de *sen embargo*, que poden ser normalizados mediante diferentes sinónimos como *porén*, *con todo*, *no entanto*... En casos coma estes, optamos en xeral por unha mesma forma estandarizadora para todos os exemplos, seleccionada de acordo con criterios de frecuencia e adecuación, pero isto non implica necesariamente que unha mesma forma desviante se normalice sempre da mesma maneira, posto que o contexto ás veces é máis axeitado para unha *target hypothesis* que para outra (por exemplo, *conlevar* estandarízase ás veces con *implicar* e outras con *supoñer*).

Noutros casos, a opcionalidade na escolla de *target hypothesis* deriva da existencia de diferentes interpretacións ao respecto do que a persoa autora dos textos quería realmente dicir. Así, por exemplo, nunha das redaccións, falando de fútbol, dise que é un «deporte con moitos afiliados», cun emprego da palabra *afiliados* semanticamente inadecuada. Estandarizamos a voz por

seguidores, pero podería ser tamén normalizada mediante *socios*, que non porta o mesmo significado ca a palabra que finalmente seleccionamos.

Ás veces o contexto pode axudar a interpretar o sentido buscado pola persoa que escribe os textos. Así, nun exemplo como o seguinte,

- (14) O consumismo e a produtividade é un dos temas máis tratados na actualidade, debido a relación de dependencia que se establece entre elas, e *como* algún cambio nunha pode repercutir na outra

podemos deducir, pola frase introducida por *debido a*, que o valor de *como* (destacado por nós en cursiva) é causal, de modo que o normalizamos con *dado que*.

As persoas usuarias do corpus deben ter sempre en mente que as propostas de estandarización realizadas son en moitas ocasións unha simple escolla de entre varias opcións posibles. En calquera caso, tal e como indicamos, procuramos a confluencia na selección mediante o establecemento previo dunha forma estandarizadora naquelas formas desviantes do corpus recorrentes.

Nos dous tipos de casos sinalados, a subxectividade afecta exclusivamente á selección da forma normalizada, pero non á anotación do tipo de desviación do estándar. Noutros casos, a escolla atangue a esta. Nun exemplo como o anterior, (14), poderíamos en principio pensar que a presenza de *a* en vez de *á*-en «debido a relación de dependendencia...») ten tres interpretacións posibles: pode verse como un problema de acentuación, como ausencia da preposición ou como ausencia de artigo.

Agora ben, nun caso como o que acabamos de presentar, non podemos obviar as características xerais das persoas que crearon os textos: estamos tratando (cando menos na inmensa maioría dos casos) con estudantes que estudan galego e en galego desde primaria e resulta pouco esperable que haxa problemas á hora de usar esa estrutura, coincidente ademais co castelán (cambiando os complementos, parece pouco probable atopar *debido iso*, sen preposición, ou *debido a motivo indicado*, sen artigo), pero si é frecuente atopar problemas na acentuación (*vid.* o capítulo de López-Sández e Lorenzo-Herrera neste volume). Particularmente, a ausencia de acento na contracción do artigo definido feminino coa preposición *a* é moi habitual (300 casos en *CORTEGAL*, computando só a contracción co artigo feminino singular).

Pero outras veces, a escolla da anotación que se debe realizar ofrece bastantes máis dificultades: o adxectivo *sana*, que se rexistra 14 veces en *CORTEGAL* e cuxa correspondente forma estándar é *sa*, pode interpretarse como unha formación non estándar do feminino de *san* (seguindo o modelo de

folgazán, folgazana) e, daquela, como unha desviación morfolóxica (M_gen_su), ou como unha desviación léxica (L_w_su), pola escolla do castelanismo *sano, sana* en vez da unidade léxica estándar do galego *san, sa*.

En casos como os indicados, sobre todo nos máis dubidosos, como o de *sana*, preséntasenos unha dobre opción metodolóxica: escoller unha soa das anotacións ou anotar a forma diverxente con todos os códigos posibles. Malia o nesgo que isto pode supoñer, inclinámonos pola primeira das opcións: escoller un único código identificador do tipo de desviación. A razón fundamental é o feito de que empregar máis dunha etiqueta suporía non poder realizar de maneira doada o cómputo de desviacións do estándar presentes no corpus, identificables e computables precisamente a través dos códigos asignados, posto que unha mesma desviación estaría anotada con varias etiquetas.

Á hora de escoller a anotación que se vai realizar entre as varias opcións posibles, empregamos como criterios fundamentais a naturalidade e frecuencia do tipo de desviación. Así, tal e como sinalamos, resultaría moi estraña a omisión de preposición ou artigo en (14), de tal xeito que etiquetamos con O_ac_om (omisión de acento gráfico). No segundo exemplo, dado que no corpus hai máis casos de *san* que de *sano*, optamos por consideralo, en xeral, un problema morfolóxico. Con todo, tamén temos en conta, como criterio secundario, o contexto: se unha persoa que utiliza *sana* emprega no mesmo texto *sano*, interpretamos *sana* como unha desviación léxica. Do que acabamos de expoñer, dedúcese que unha mesma forma desviante pode ter ocasionalmente dúas anotacións diferentes en textos distintos, condicionadas pola información que se extrae do contexto.

A subxectividade da anotación vai implícita tamén na aplicación xeral dalgúns etiquetas concretas. Por exemplo, a persoa anotadora debe decidir cando a selección inadecuada de xénero responde a un problema gramatical de concordancia (G_gen_su) e cando obedece a unha atribución de xénero non estándar a unha unidade léxica, en cuxo caso, a etiqueta sería léxica (L_gen_su). Así, por exemplo, consideramos que en «a leite francesa» hai un problema léxico, anotado sobre *leite*, posto que entendemos que se lle atribúe a esta palabra xénero feminino. Pola contra, en «a cultura xeralmente abre máis portas e non termina sendo tóxico» a anotación realízase sobre *tóxico* e non sobre *cultura*, dado que se considera que hai un problema de concordancia do adxectivo e non de asignación a *cultura* de xénero masculino. A razón do distinto tratamento obedece a criterios de frecuencia (hai en CORTEGAL 15 exemplos de uso en feminino de *leite*) e á existencia

dalgũa razón que xustifica a atribución dun xénero non estándar (o cognado do castelán *leche* é masculino). Decisións moi similares afectan á distinción entre erros ortográficos de acentuación (por exemplo, *ademaís*, considerado como un erro na colocación do acento gráfico: O_ac_wp) e erros léxicos de selección non estándar da sílaba tónica (por exemplo, *élite*, fronte ao galego estándar *elite*, considerado como un erro léxico: L_ac_su). As razóns son do mesmo tipo ca as indicadas no caso anterior (hai 7 exemplos de *élite* e existe o cognado español *élite*). Pode atoparse máis información sobre estas cuestións e sobre outras decisións do mesmo tipo no capítulo de Álvarez de la Granja neste volume.

Finalmente, debemos mencionar aqueles casos en que a opcionalidade da estandarización radica na posibilidade de normalizar elementos diferentes, o que pode implicar anotacións de distinto tipo.

Así, por exemplo, nun fragmento como este

- (15) Pola contra noutros países que non teñen tantos recursos como por exemplo os países máis pobres de África, non producen nin consumen como os países europeos

poderíamos optar hipoteticamente por dúas estandarizacións (incluímos só as modificacións que afectan ao que nos interesa ilustrar):

- (16) Pola contra, *outros* países que non teñen tantos recursos como por exemplo os países máis pobres de África, non producen nin consumen como os países europeos
- (17) Pola contra noutros países que non teñen tantos recursos como, por exemplo, os países máis pobres de África, non *se produce* nin *se consume* como *nos* países europeos

No primeiro caso, o adxunto inicial «pasa a funcionar» na nova estrutura como suxeito. No segundo caso, modificamos a segunda parte do fragmento, mantendo o adxunto como tal e, dada a ausencia de suxeito, empregando unha estrutura impersoal.

En situacións deste tipo, o criterio xeral para a escolla da estandarización é o de optar pola forma normalizada máis sinxela e máis próxima ao texto da/do estudante, que neste caso corresponde a (16).

Así pois, e a maneira de conclusión do exposto, aínda que a subxectividade na estandarización e etiquetaxe é inevitable, procuramos utilizar algúns criterios que aseguren a maior homoxeneidade posible: por un lado, descartamos sempre aquelas interpretacións dunha desviación que son posibles pero estrañas ou forzadas. Buscamos sempre a interpretación máis natural

tendo en mente as características xerais das persoas que elaboraron os textos, as cales, teñan o galego como L1 ou L2, estudan e coñecen a lingua, cando menos na gran maioría dos casos, desde a educación primaria. Ademais, o criterio da frecuencia é fundamental, tanto para a escolla das formas normalizadas como para determinar a anotación que se realiza. Tamén temos en conta a proximidade ao texto orixinal e a información que este nos pode proporcionar, tanto á hora de estandarizar como de anotar. Debemos indicar, ademais, que a posibilidade que nos ofrece *TEITOK* de editar conxuntamente varias formas desviantes (por exemplo, podemos estandarizar e etiquetar conxuntamente todos os exemplos de *aunque* presentes en *CORTEGAL*), non só facilita o proceso, senón que contribúe a diminuír o número de erros e a lograr maior coherencia.

Os criterios expostos están recollidos no manual de estandarización e etiquetaxe, xa mencionado en §2.5.4, que estivo a disposición das persoas responsables desas dúas tarefas ao longo de todo o proceso. Antes da aplicación das indicacións do manual no corpus, realizáronse probas co obxectivo de adestrar as persoas encargadas das dúas tarefas, todas elas lingüistas bilingües en galego e castelán. O manual contén os criterios xerais xa mencionados, así como os códigos empregados na anotación, con exemplos, explicacións e observacións adicionais sobre o seu emprego, e tres apéndices para a estandarización de aspectos especialmente problemáticos: uso dos signos de puntuación, uso do xerundio e uso da preposición con obxecto directo.

A partir da análise dunha mostra de textos de *CORTEGAL* e das propostas realizadas para outros corpus con anotación informatizada de erros elaborouse unha versión piloto para a estandarización e etiquetaxe, que se testou con outra pequena selección de textos. Tendo en conta os problemas detectados, elaborouse unha segunda versión do manual, con algunhas modificacións na codificación provisoria e co engadido de novas explicacións e aclaracións. Esta nova versión avaliouuse á vista das dificultades de aplicación e do acordo intra- e interanotadoras/es tanto para a estandarización como para a etiquetaxe dun conxunto de textos, resultando unha terceira versión, coa que se deu por rematada a pilotaxe. Con todo, esta versión sufriu tamén varias modificacións ao longo do proceso de normalización e codificación dos 1000 textos, dada a necesidade de realizar algúns axustes e aclaracións. Para asegurar a maior homoxeneidade posible, a estandarización e etiquetaxe de todos os textos foi revisada por unha das persoas responsables do corpus, a cal, unha vez rematado o proceso, tamén levou a cabo a revisión da aplicación coherente dos códigos a través da análise das liñas de concordancia proporcionadas polo buscador. Desta revisión derivaron novos cambios, para

lograr unha maior consistencia na etiquetaxe, que conduciron á versión final do manual, que é a que se encontra dispoñible, como xa foi indicado, na páxina do proxecto. Por suposto, despois de cada revisión do manual, levouse a cabo tamén a revisión da estandarización e da etiquetaxe realizadas ata ese momento para aplicar os novos criterios, asegurando deste xeito a maior coherencia posible ao longo de todo o corpus.

A pesar das continuas revisións, son inevitables algunhas inconsistencias, que corriximos a medida que detectamos. O corpus está aberto tamén á participación das persoas usuarias, que poden facer chegar as súas suxestións e indicacións ás/aos responsables de *CORTEGAL*. En calquera caso, tales inconsistencias non deben confundirse coas interpretacións, inevitables na estandarización e etiquetaxe dos corpus de aprendentes: non podemos considerar que «the error annotation which is present in a corpus is the ‘truth’ or ‘correct analysis’ instead of just one among many interpretations» (Lüdeling e Hirschmann, 2015, p. 142).

2.6. Asignación de lemas e categorías gramaticais (estándar e orixinais)

A lematización realízase automaticamente en *TEITOK* a través da versión 4.2 de *FreeLing* en lingua galega. O proceso de lematización con *FreeLing* implica a asignación de lema estándar e categoría gramatical a todos os *tokens*, incluídos os signos de puntuación (no caso das contraccións e verbos con pronomes enclíticos, só aos *dtokens*, vid. §2.2). A anotación morfosintáctica emprega etiquetas *EAGLES* (Leech e Wilson, 1996).

A lematización realízase unha vez normalizadas e etiquetadas as formas non estándar e sempre sobre a versión lexicamente estandarizada, que implica tamén a estandarización ortográfica e morfolóxica, tal e como vimos en §2.5.3. Escóllese a dimensión léxica, pois é o nivel máis próximo ao texto orixinal que asegura a identificación e lematización adecuada dos *tokens* por parte de *FreeLing*.

Deste xeito, a lematización dun texto como (18), en que se destacan en cursiva as formas que deben ser estandarizadas na dimensión ortográfica (*conflictos*), na morfolóxica (*corporales*) e na léxica (*sen embargo, tareas*),

- (18) Na época de adolescencia é unha época chea de *conflictos* xa que é unha etapa dura nosas vidas, e cando somos tan novos e inexpertos necesitamos o apoio de figuras paternas e maternas, *sen embargo* non é unha *tarea* doada xa que é a idade con máis cambios de humor e incompreensión por parte dos nenos xa que está chea de cambios, tanto *corporales* como na mentalidade

realízase sobre (19), unha vez realizadas as correspondentes modificacións, posto que formas como *corporales*, *tarea* ou *conflictos* non son recoñecidas por *FreeLing*.

- (19) Na época de adolescencia é unha época chea de *conflictos* xa que é unha etapa dura nosas vidas, e cando somos tan novos e inexpertos necesitamos o apoio de figuras paternas e maternas, *porén* non é unha *tarefa* doada xa que é a idade con máis cambios de humor e incompreensión por parte dos nenos xa que está chea de cambios, tanto *corpormais* como na mentalidade

Pola contra, o lematizador non alcanza as estandarizacións que se realizan no nivel gramatical, semántico ou discursivo. Así, por exemplo, no texto anterior substitúese no nivel discursivo a coma que precede a *sen embargo* / *porén* por un punto, pero o lema e categoría gramaticais asignados son «,» e «Fc» e non «.» e «Fp». De igual xeito, a contracción *das*, incluída no nivel gramatical despois de «unha etapa dura», non é obxecto de lematización, como tampouco o é ningún outro elemento incorporado no texto polas persoas encargadas da estandarización cando hai algunha omisión.

Un aspecto metodolóxico importante de *CORTEGAL* é a asignación manual, na maior parte das formas estandarizadas lexicamente, dun segundo lema, chamado lema orixinal e, nunhas poucas destas formas, dunha segunda categoría gramatical, denominada clase gramatical orixinal. Así, a forma *tarea* que figura en (18) recibe dous lemas: o lema estándar *tarefa*, asignado automaticamente mediante *FreeLing*, e o lema orixinal *tarea*, asignado manualmente. O lema orixinal é a forma de cita que representa a voz escrita pola/polo estudante (unha vez estandarizada, se é o caso, ortográfica e morfoloxicamente), así como calquera outra variante flexiva que poida haber no corpus. Así, outros exemplos de lemas orixinais introducidos en *CORTEGAL* son *sen embargo*, *abuelo* (común a *abuela*, *abuelo* e *abuelos*, as tres formas recollidas no corpus), *conlevar* (representante de *conleva*, *conlevan*, *conlevar*, *conlevaron* e *conlevou*), *novidoso* (asignado a *novidoso*, *novidosa*, *novidosos*, *novidosas* e ao lapsus ortográfico *novidosos*), *cachivache* (atribuído a *cachibaches*) etc. No caso de que non se atribúa ningún lema orixinal manualmente, este coincide por defecto co lema estándar. Tal e como se indicou, o lema orixinal só se lles asigna manualmente ás formas que son estandarizadas no nivel léxico, concretamente a aquelas que portan un código L_w_su (substitución dunha unidade léxica estándar por unha non estándar, con diferenzas que van máis alá da acentuación de intensidade ou do xénero, como *abuela* por *avoa*) ou un código L_ac_su (substitución dunha unidade léxica estándar por unha non estándar, con diferenzas de acentua-

ción de intensidade, como *élite*). Deste xeito, nas formas etiquetadas con códigos diferentes a L_w_su e L_ac_su, así como nas voces que non reciben ningún tipo de modificación, o lema orixinal coincide co estándar.

A categoría gramatical orixinal introdúcese tamén manualmente cando existe algunha diferenza entre a clase ou subclase de palabras da forma escrita pola/polo estudante e da forma estandarizada no nivel léxico. O exemplo máis claro é o das voces etiquetadas con L_gen_su, en que se produce a asignación dun xénero non estándar a determinadas palabras, tal e como vimos en §2.5.5 e §2.5.6. Deste xeito, a palabra *leite* na secuencia «a leite» ten atribuído como clase de palabra estándar NCMS000 (substantivo común masculino singular), pero como clase de palabra orixinal NCFS000 (substantivo común feminino singular). Máis alá das formas etiquetadas con L_gen_su, podemos atopar outros casos de asignación manual de clase de palabra orixinal: nótese, por exemplo, a diferenza categorial entre o lema orixinal *aunque*, conxunción, e o lema estándar *aínda que*, adverbio + conxunción (*vid.* o indicado ao final deste apartado sobre o tratamento das expresións complexas). Como sucedía cos lemas, a asignación de clase de palabra orixinal só se realiza nas palabras estandarizadas no nivel léxico, de maneira que as estandarizacións que se realicen no nivel gramatical, semántico ou discursivo e que impliquen cambio categorial non implican a atribución manual dunha clase de palabra orixinal. Se non se atribúe ningunha categoría gramatical manualmente, esta coincide coa estándar, de xeito que naqueles casos en que a estandarización non implica mudanza categorial, así como nos *tokens* que non se modifican, a clase de palabra orixinal coincide coa estándar, atribuída por *FreeLing*.

A asignación de lema orixinal é especialmente útil na análise dos datos do corpus, sobre todo cando interesa realizar consultas sobre frecuencia tendo os lemas como criterio de agrupamento. Mediante as consultas sobre frecuencia que ofrece o buscador de *TEITOK*, é posible obter a distribución dalgún elemento (por exemplo, un código de anotación das formas non estándar) organizado de acordo con algún criterio (por exemplo, lemas). Así, por exemplo, se interesa obter unha listaxe lematizada das formas codificadas con D_reg_su, pode realizarse unha busca das formas etiquetadas con este código e posteriormente seleccionar como criterio organizador os lemas estándar ou ben os lemas orixinais. Deste xeito, obtense unha táboa coa listaxe de lemas, tal e como se ve na Figura 24 (ofrecemos un fragmento).

FIGURA 24. Frecuencia de D_reg_su por lema orixinal (fragmento da táboa xerada por TEITOK)

Distribución no corpus

Consulta	Tipo de desviación do estándar = (.*)?D_reg_su(,.*)?
Criterio de agrupamento	Lema orixinal
Total	274
Tamaño de referencia	274617

Gráfica: | Contar: | Descargar:

Lema orixinal	Frecuencia	Porcentaxe (%)	Porcentaxe (%)
cousa	54	196.64	19.71
ser	31	112.88	11.31
estar	18	65.55	6.57
tele	18	65.55	6.57
isto	14	50.98	5.11
saír	13	47.34	4.74
que	10	36.41	3.65
iso	9	32.77	3.28
coller	8	29.13	2.92

Agora ben, se se seleccionan os lemas orixinais obteremos unha listaxe máis próxima ao texto da/do estudante (por exemplo, teremos nesa listaxe *asqueroso*, correspondente á voz *asquerosa* escrita pola/o estudante). Se non houberse lema orixinal e só contásemos coa posibilidade de escoller como criterio organizador o lema estándar, na listaxe aparecería *noxento*, pois o castelanismo *asquerosa* foi estandarizado no nivel léxico mediante *noxenta*, de xeito que foi esta a voz que se lematizou.

O interese do lema orixinal é obvio cando o que se pretende estudar son as formas lexicamente non estándar (por exemplo, as etiquetadas con L_w_su), porque só deste xeito podemos obter unha listaxe lematizada das formas non estándar escritas polo alumnado, tal e como vemos na Figura 25 (fragmento). Por suposto, se só tivéssemos lemas estándar, isto non sería posible, porque a táboa só incluíría estes últimos (Figura 26, fragmento).

FIGURA 25. Frecuencia de L_w_su por lema orixinal (fragmento da táboa xerada por TEITOK)

Distribución no corpus

Consulta	Tipo de desviación do estándar = (.*;)?L_w_su(,.*)?
Criterio de agrupamento	Lema orixinal
Total	2868
Tamaño de referencia	274617

Gráfica: | Contar: | Descargar:

Lema orixinal	Frecuencia	Media	Desviación estándar
plato	103	375.07	3.59
disfrutar	92	335.01	3.21
convertir	67	243.98	2.34
precio	55	200.28	1.92
sen+embargo	52	189.35	1.81
si	48	174.79	1.67
receta	38	138.37	1.32
basar	35	127.45	1.22

FIGURA 26. Frecuencia de L_w_su por lema estándar (fragmento da táboa xerada por TEITOK)

Distribución no corpus

Consulta	Tipo de desviación do estándar = (.*;)?L_w_su(,.*)?
Criterio de agrupamento	Lema estándar
Total	2868
Tamaño de referencia	274617

Gráfica: | Contar: | Descargar:

Lema estándar	Frecuencia	Media	Desviación estándar
prato	103	375.07	3.59
gozar	92	335.01	3.21
converter	67	243.98	2.34
porén	59	214.84	2.06
prezo	55	200.28	1.92
calquera	54	196.64	1.88
se	48	174.79	1.67
receita	38	138.37	1.32
custo	36	131.09	1.26
basear	35	127.45	1.22

Por outro lado, e dado que o lematizador empregado non identifica como unidades as expresións complexas lexicalizadas, cada un dos seus constituíntes é lematizado de maneira independente. Así, por exemplo, o conector *pola contra* non constitúe unha unidade recoñecida como tal por *FreeLing*, de maneira que cada constituínte ten un lema e unha categoría gramatical asignados de maneira independente (preposición *por*, artigo determinado *o*, substantivo *contra*). Unicamente no caso de que sexa necesaria unha estandarización conxunta dunha expresión, levamos a cabo a súa unificación manual mediante a función *merge* que ofrece *TEITOK*, tal e como indicamos en §2.5.4, pero neste caso mantemos tamén a asignación de lemas e categorías gramaticais para cada compoñente, separados desta volta mediante o signo «+»: así, por exemplo, a locución *encher o buche*, que debe unificarse, posto que é substituída por *saciarse* no nivel discursivo ao non pertencer a un rexistro formal, ten como lema orixinal *encher + o + buche*, como lema estándar *encher + o + boche* e como clase de palabras (estándar e orixinal, pois non varía) VMN0000+DAoMS0+NCMS000 (infinitivo + artigo determinado masculino singular + nome común masculino singular).

2.7. Clasificación dos conectores que vinculan enunciados

TEITOK é unha plataforma flexible, que permite anotacións de moi diferente tipo e a incorporación destas en calquera momento. Consideramos que, dado que o estudo dos aspectos discursivos en lingua galega non recibiu ata o de agora demasiada atención, sería interesante a anotación no corpus dalgúns elementos con relevancia no nivel textual (máis alá das desviacións) para promover análises neste ámbito. Con todo, a complexidade e a esixencia do resto das anotacións determinou que esta tarefa se limitase ata o de agora a unha anotación sinxela daqueles conectores que teñen como función a vinculación entre enunciados (excluíndo vínculos temporais relativos á ordenación cronolóxica dos feitos na realidade⁹), aínda que a nosa intención é engadir no futuro novas anotacións discursivas.

⁹ Así, por exemplo, non se anotaría o conector *despois*, nun uso como o que atopamos en «Os anos de posguerra e de franquismo foron anos de pobreza; a xente comía o que se podía permitir cociñar na súa casa. *Despois*, coa recuperación económica, a xente comezou a frecuentar máis os restaurantes». Si se anotaría *despois*, pola contra, nestoutro fragmento, posto que neste caso o conector serve para organizar as ideas do discurso: «Polo tanto hai distintos sectores de produción. Por exemplo, no sector de produción téxtil, vese condicionado pola moda de cada época. A nós tocounos vivir nunha época no que a moda ten un papel moi activo na nosa sociedade. Entón unha cousa dependerá da outra, dependendo das tendencias de cada momento a produción será distinta e as persoas beneficiadas tamén. *Despois* hai sectores como os da alimentación, que non dependen do mesmo xeito da poboación, porque todo o mundo ten que comer, aínda que sexa o máis básico».

De acordo co indicado, anótanse en *CORTEGAL* os conectores que serven para unir enunciados, pero non os que establecen vínculos intraoracionais. Deste xeito, son etiquetados os elementos destacados en (20) e (21), pero non en (22) e (23):

- (20) O consumo consiste na acción que realizamos todos de adquirir algo a cambio dun determinado prezo (capital). *Por outra banda*, ese algo que adquirimos é o produto
- (21) Sen consumo, non hai produción, porque a demanda é inexistente, e se non hai produción, pois non pode haber consumo.
Pero o que parece fácil, se o analizamos profundamente, non o é
- (22) Como se relata no texto, o consumo é unha necesidade e eso é moi certo, posto que é necesario para mercar a comida de cada día, *pero por outra banda* tamén di que serve para obter a máxima satisfacción
- (23) Moitos dirán que tan só é unha moda, *pero* é necesario sinalar que a cociña e a gastronomía cambiáronlle a vida a moitas persoas

Os conectores textuais foron clasificados en 13 grupos de acordo co valor que achegan. Os códigos, valores e exemplos amósanse na Táboa 5:

TÁBOA 5. Códigos empregados para a anotación dos conectores que vinculan enunciados

Código	Valor	Exemplos
ad	Introducen un elemento que se suma ao dito anteriormente na argumentación, mesmo reforzándoo.	<i>ademais, e, é máis, tamén...</i>
caus	Introducen a causa do expresado no enunciado ou enunciados previos	<i>ao fin e ao cabo, pois, porque...</i>
coment	Introducen un comentario sobre o expresado anteriormente	<i>ben, pois, pois ben...</i>
concl	Introducen o resultado ao que se chega a partir do exposto previamente ou ben o seu resumo	<i>en conclusión, en definitiva, en resumo...</i>
confirm	Introducen un argumento, explicación... que serve para confirmar o que se acaba de dicir	<i>efectivamente, en efecto, si</i>
cons	Introducen a consecuencia do expresado no enunciado ou enunciados previos	<i>así, entón, por conseguinte...</i>
contraarg	Introducen un argumento ou explicación que se contrapón dalgún xeito ao que se acaba de indicar	<i>aínda así, agora ben, en cambio...</i>
disjunc	Introducen unha alternativa ao que se acaba de expoñer	<i>ou</i>
exempl	Introducen un exemplo ou concreción do que se acaba de indicar	<i>así, concretamente, por exemplo...</i>

TÁBOA 5. Continuación

fin	Introducen a finalidade do expresado no enunciado ou enunciados previos	<i>para</i>
ord_sp	Organizan o discurso distribuindo os feitos expostos en diferentes apartados mediante elementos con valor espacial	<i>por outra banda, por unha parte, por un lado...</i>
ord_t	Organizan o discurso indicando a posición que ocupa cada elemento no proceso argumentativo ou expositivo do texto mediante elementos con valor temporal	<i>finalmente, por último, primeira-mente...</i>
reform	Explican ou presentan doutra forma o que se acaba de expoñer	<i>é dicir, explícome, isto é</i>

Para levar a cabo a anotación destes conectores da maneira máis coherente posible, foi necesario tomar unha serie de decisións relativas tanto á delimitación da unidade «enunciado» como do concepto «conector». Estas decisións, que por razóns de espazo non poden ser comentadas aquí, están recollidas nun manual de anotación, que está dispoñible na páxina do proxecto (<http://ilg.usc.gal/cortegal/downloads/conectores.pdf>). Unicamente nos gustaría sinalar o feito de que a clasificación dos conectores se realiza atendendo ao valor que se lles atribúe por defecto¹⁰, con independencia de que, no contexto en que figuran, o seu uso non sexa adecuado. Así, por exemplo, o conector *pero* que rexistramos no fragmento recollido en (24) é clasificado como contraargumentativo, por moito que no contexto non pareza haber ningún tipo de contraargumentación (polo que se anota con S_w_su e se substitúe por *e*).

- (24) Na actualidade a gastronomía e a cociña acadaron moita popularidade. A gastronomía dos diferentes países é moi abundante pero tamén moi diferentes entre si. *Pero* na actualidade son un factor clave á hora de visitar determinados países

Constitúen excepción os casos en que existen razóns para supoñer que estamos, non ante un mal uso, senón ante unha atribución errónea de significado, como sucede con *porén*, utilizado en varios textos con valor consecutivo, de tal xeito que lle atribuímos en tales textos o código *cons*. En (25) atopamos un exemplo deste tipo:

- (25) Se a sociedade deixase de consumir produtos, as fábricas deixarían de producir e produciríase unha xigantesca crise económica, debido a que moitas persoas quedarían sen traballo, non recibirían un soldo, e non poderían consumir. *Porén*, Podemos dicir que é «unha pescadilla que se morde a cola»

¹⁰ Por suposto, se o conector é polisémico, como sucede, por exemplo, con *así*, analizamos o contexto para atribuírlle o código que lle corresponde.

Finalmente, gustaríanos indicar que, no caso dos conectores pluriverbais (*así pois, por outra banda, sen embargo...*), o código asígnaselle a un dos elementos da unidade (aquele que parece achegar unha maior carga semántica; nos casos anteriores *pois, banda e embargo*), pero non levamos a cabo ningún tipo de fusión mediante *merge*, pois, como indicamos en §2.5.4, esta opción dá lugar a algúns problemas de recuperación e de etiquetaxe.

3. PRINCIPAIS RESULTADOS

3.1. Caracterización xeral dos textos. Análise cuantitativa

O número total de palabras dos textos é de 224.424, de acordo cos datos que ofrece *Dcontado*, que tal e como indicamos en §2.6, non computa os signos de puntuación e contabiliza as contraccións e as combinacións de verbo e pronome enclítico como unha soa unidade. Dado que o corpus está conformado por 1000 textos, a media de palabras por texto é de 224,42. Do total de palabras, 202.645 corresponden aos exames da convocatoria de xuño e 21.779 aos de setembro, de tal xeito que os 898 exames da primeira convocatoria teñen unha media de 225,66 palabras por texto e os 102 exames da segunda convocatoria unha media de 213,52, inferior á anterior en algo máis de 12 palabras. O texto cun menor número de palabras contén 51 (semella un texto que quedou sen rematar) e o máis longo ten 613. A desviación estándar é de 52,74.

Con respecto ao número de lemas ou palabras distintas, establecido tamén de acordo cos datos obtidos de *Dcontado*, este é de 113.348, de xeito que temos unha media de 113,35 lemas por texto. Os lemas distribúense entre 102.593 en xuño e 10.755 en setembro, o que supón unha media de 114,25 lemas en xuño e 105,44 en setembro, cunha diferenza de 8,81. O rango sitúase entre un mínimo de 37 e un máximo de 270 lemas, nos mesmos textos que servían para establecer o rango de palabras. A desviación estándar é de 22,69.

Para obter a densidade léxica, dividimos o número de lemas entre o número de palabras¹¹. A densidade léxica media é de 0,51, cunha pequena diferenza entre os textos de xuño (0,51) e os de setembro (0,50), malia que, en principio, a razón entre *types* e *tokens* adoita diminuír a medida que aumenta a lonxitude do texto (Capsada Blanch e Torruella Casañas, 2017, p. 351) (e lembremos que os textos de xuño son máis longos ca os de setembro).

¹¹ Somos conscientes de que esta medida de densidade léxica, coñecida como TTR (*vid.* Capsada Blanch e Torruella Casañas, 2017), é moi simple e non ten en conta aspectos relevantes, como a extensión dos textos, pero razóns prácticas determinaron a escolla deste sistema de cálculo. En calquera caso, os datos e os textos están á disposición de calquera persoa que desexe realizar unha medición máis axustada.

O rango sitúase entre 0,35 e 0,73 (no texto máis breve, de 51 palabras e 37 lemas), sendo a desviación estándar de 0,05.

Lémbrese en calquera caso as precaucións indicadas en §2.6 sobre o cómputo de lemas e, daquela, da densidade léxica.

No que se refire ao número de enunciados, *DContado* contabiliza como tales as unidades que van de punto a punto. A media no total dos textos é de 9,03, sendo a diferenza entre xuño e setembro de máis dun enunciado (9,15 en xuño e 8,06 en setembro). O rango sitúase entre 2 enunciados en 3 textos e 29 nun (non se trata do texto máis longo, senón dun de 356 palabras). A desviación estándar é de 3,23.

Con respecto á media de palabras por enunciado, esta é de 26,96 e a diferenza entre xuño (26,66) e setembro (29,51) é de case tres palabras, pero desta volta a cifra máis alta corresponde á segunda convocatoria. O rango establécese entre 10,05 e 98,50 (nun dos textos con só dous enunciados), sendo a desviación estándar de 8,88.

A media de palabras do enunciado máis longo é de 47,51, e é tamén en setembro onde atopamos unha media máis alta, con 48,59 fronte a 47,39 en xuño, cunha diferenza de algo máis dunha palabra. O rango vai de 17 a 146 e a desviación estándar é de 15,91. E, con respecto ao enunciado máis curto, a media sitúase en 12,28, con 11,94 palabras en xuño e 15,32 en setembro. O rango vai de 1 (en 7 textos, estando o enunciado constituído en varios deles pola palabra *non*) a 93 nun texto. A desviación estándar é de 6,49.

En relación co número de párrafos, a media é de 4,30, algo superior en xuño (4,32) ca en setembro (4,07). Un total de 44 textos están constituídos por un só párrafo, mentres que o texto con maior número de párrafos ten 12. A desviación estándar, pola súa parte, é de 1,60.

Finalmente, a media de enunciados por párrafo é de 2,41, sendo algo máis alta na primeira convocatoria (2,42) ca na segunda (2,33). O rango sitúase entre un só enunciado por párrafo en 40 textos e 11 en 5 textos, sendo a desviación estándar de 1,39.

Os datos presentados recóllense na Táboa 6:

TÁBOA 6. Datos cuantitativos sobre os textos de CORTEGAL

Media total	Media xuño	Media setembro	Desviación estándar	Rango
<i>Palabras</i>				
224,42	225,66	213,52	52,74	51-613
<i>Lemas</i>				
113,35	114,25	105,44	22,69	37-270
<i>Densidade léxica</i>				
0,51	0,51	0,50	0,05	0,35-0,73
<i>Enunciados</i>				
9,03	9,15	8,06	3,23	2-29
<i>Palabras por enunciado</i>				
26,96	26,66	29,51	8,88	10,05-98,50
<i>Palabras do enunciado máis longo</i>				
47,51	47,39	48,59	15,91	17-146
<i>Palabras do enunciado máis curto</i>				
12,28	11,94	15,32	6,49	1-93
<i>Parágrafos</i>				
4,30	4,32	4,07	1,60	1-12
<i>Enunciados por parágrafo</i>				
2,41	2,42	2,33	1,39	1-11

Tal e como se pode observar na táboa, os textos da convocatoria de xuño, que, de acordo co que vimos en §2.1, teñen unhas cualificacións considerablemente máis altas ca os da convocatoria de setembro (6,93 fronte a 5,21), son textos algo máis longos ca estes últimos. O tamaño medio dos textos (225,67 palabras) sitúase, curiosamente, na media dos dous extremos indicados na proba (solicítanse textos de entre 200 e 250 palabras). O alumnado de setembro fai menos enunciados pero algo máis longos, de 29,51 palabras, fronte ás 26,66 do alumnado de xuño, e estes enunciados distribúense nunha media total de 4,30 parágrafos e 2,41 enunciados por parágrafo, sendo ambas as medias bastante similares nas dúas convocatorias. Con todo, o rango da media de enunciados por parágrafo que máis se repite é o que se sitúa entre 1 e 1,99 (con 434 textos), de tal xeito que un 43% dos textos da primeira convocatoria teñen unha media de enunciados por parágrafo inferior a 2 e esa porcentaxe elévase ao 48% na convocatoria de setembro. De feito, o 7,84% dos textos de setembro teñen un único enunciado por parágrafo, mentres que en xuño esa porcentaxe baixa ao 3,56%.

3.2. Tokens e clases de palabras

O número total de *tokens* que contabiliza o corpus é de 274.617, pero debe terse en conta que *TEITOK* computa tamén as formas engadidas polas persoas editoras no proceso de estandarización. Se nos limitamos ás formas que escribiu a/o estudante, que son as únicas que se lematizan e as únicas ás que se lles asigna categoría gramatical, temos un total de 268.595 *tokens*, cuxa distribución por categorías se recolle na Táboa 7. Debe sinalarse que, cando creamos un *token* complexo mediante *merge*, *TEITOK* compútao na clase do seu primeiro compoñente, de tal xeito que optamos por crear a categoría «Expresión complexa», e contabilizar conxuntamente todas estas formas constituídas por varios elementos, restándoas do cómputo do compoñente inicial¹².

TÁBOA 7. Distribución dos *tokens* en clases de palabras

Categoría	Número de <i>tokens</i>	Porcentaxe sobre o total
Substantivo	52406	19,51%
Determinante	44452	16,55%
Verbo	40936	15,24%
Preposición	35337	13,16%
Signo de puntuación	25004	9,31%
Conxunción	19420	7,23%
Pronome	18292	6,81%
Adxectivo	15945	5,94%
Adverbio	15896	5,92%
Numeral	595	0,22%
Expresión complexa	302	0,11%
Interxección	10	0,00%
<i>Total</i>	<i>268.595</i>	

3.3. As formas non estándar

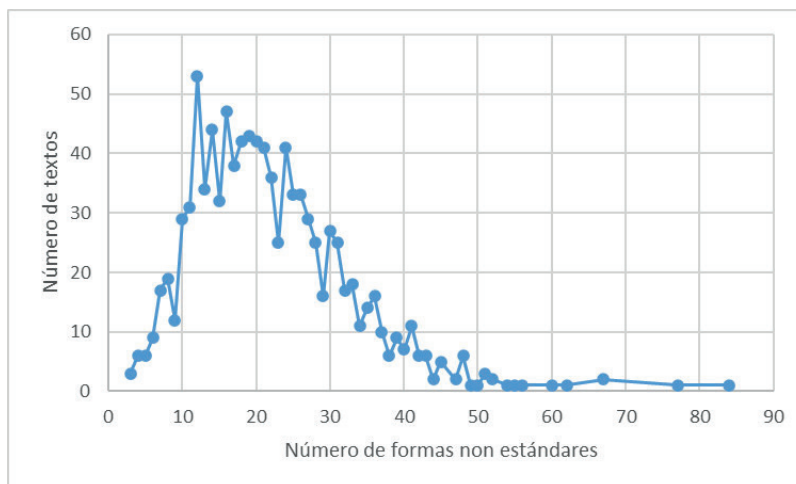
Tal e como indicamos en §2.5.4, asignámoslles a todas as formas non estándar un código que describe o tipo de desviación e, nalgúns casos, unha

¹² Por tal motivo, nalgúns categorías, a cifra que ofrecemos non coincide exactamente coa que indica *TEITOK*.

segunda etiqueta que indica a súa orixe. Centrarémonos en primeiro lugar na codificación descritiva, á que corresponden un total de 23.169 etiquetas. Isto quere dicir que a media de códigos que sinalan algún tipo de desviación do estándar é de 23,17 por texto. A cifra total e a media indicadas non coinciden exactamente co número total de *tokens* anotados no corpus e coa media de *tokens* etiquetados por texto, respectivamente, posto que a unha mesma forma poden asignárselle varios códigos e porque en varias ocasións os códigos non etiquetan formas do corpus, senón que identifican omisións de palabras, que nós introducimos en *CORTEGAL*. Se consideramos os *tokens* anotados así como os creados por nós, asociados a códigos de desviación do estándar, o número total de formas anotadas é de 21.940, cifra á que hai que sumar as 513 secuencias ou palabras anotadas en *stand-off*, o que dá como resultado un total de 22.453 *tokens* ou secuencias con códigos de forma non estándar e unha media de 22,45 por texto. Se comparamos os datos das convocatorias de xuño e de setembro, observamos que a media é máis alta nesta segunda oportunidade, o que resulta coherente co feito de que as cualificacións desta convocatoria sexan máis baixas (*vid.* §2.1). Así, a media de formas con códigos non estándar en xuño é de 22,07 e en setembro de 25,82.

A desviación estándar no conxunto dos textos (sen ter en conta as anotacións en *stand-off*, que son computadas á parte en *TEITOK*) é de 10,43 e o rango sitúase entre 84 formas anotadas a nivel de *token* nun texto e 3 anotacións en *stand-off* (así pois, un total de 87) e 3 formas en 3 redaccións, de xeito que ningún dos textos está libre de anotacións deste tipo (en xuño o rango vai de 77 nun texto máis 3 anotacións en *stand-off*, isto é, un total de 80, a 3 formas anotadas en 3 textos, mentres que en setembro vai de 87 a 7, nun texto en cada caso). Na Gráfica 6 observamos a distribución das formas non estándar entre os textos, deixando de novo a un lado as anotacións *stand-off*. Podemos ver como as cifras máis recorrentes se sitúan entre as 10 e as 30 formas non estándar (as cifras que máis se repiten son 12 e 16, en 53 e 47 textos respectivamente). Así mesmo, comprobamos que son escasos tanto os textos con poucos *tokens* anotados (tan só 15 textos con 5 formas ou menos) coma aqueles con 50 ou máis (tamén 15 textos).

GRÁFICA 6. Distribución das formas non estándar nos textos de CORTEGAL



Por outra parte, a distribución dos códigos de formas non estándar entre os diferentes niveis (ortográfico, morfolóxico, léxico, gramatical, semántico e discursivo) é a que se ofrece na Táboa 8¹³.

TÁBOA 8. Códigos descritivos das formas non estándar por nivel

Nivel	Número de códigos	Porcentaxe
Discursivo	7366	31,79%
Ortográfico	5648	24,38%
Gramatical	3958	17,08%
Léxico	3049	13,16%
Semántico	2004	8,65%
Morfolóxico	1144	4,94%
<i>Total</i>	23.169	

Tal e como se pode observar na Táboa 8, as etiquetas máis frecuentes son as do ámbito discursivo, seguidas polas do nivel ortográfico. Unhas e outras supoñen máis do 50% das etiquetas asignadas. As menos frecuentes son as do ámbito morfolóxico, relativas á flexión, cun 5% dos códigos atribuídos.

¹³ Uns poucos códigos do ámbito léxico e do ortográfico poden aparecer máis dunha vez asociados ao mesmo *token* (por exemplo, *O_cons_su*, cando se produce unha selección inadecuada de dúas consoantes, como en *susesibamente*, ou dous códigos *L_w_su* cando un calco parcial dunha unidade complexa presenta unha transferencia directa do español no seu interior, como ocorre con *perrito quente*). Neste caso, *TEITOK* só computa un código, co cal estas poucas etiquetas duplicadas foron computadas manualmente.

A preponderancia dos códigos no nivel discursivo explícase pola existencia de moitos problemas relativos ao uso dos signos de puntuación, cun total de 5979 etiquetas, dos cales máis do 60% corresponden a omisións, particularmente de comas. Na Táboa 9 ofrécense os 10 códigos con máis de 500 asignacións, onde se pode comprobar como o código que identifica estas omisións de signos de puntuación, D_pm_om, é o máis frecuente (15,84% do total de etiquetas do corpus). Na táboa tamén podemos ver como entre as etiquetas máis frecuentes existen códigos pertencentes aos seis niveis discursivos:

TÁBOA 9. Códigos descritivos con máis de 500 ocorrencias

Código	Valor	Número de códigos asignados
D_pm_om	Omisión dun signo de puntuación	3670
L_w_su	Selección dunha unidade léxica non estándar con diferenzas que van máis aló do xénero ou da acentuación	2887
O_ac_om	Omisión de acento gráfico	2363
S_w_su	Selección dunha unidade léxica inadecuada desde o punto de vista semántico ou combinatorio	1391
D_pm_su	Selección inadecuada dun signo de puntuación	1223
D_pm_ad	Adición dun signo de puntuación	919
G_num_su	Selección de número non estándar	854
O_cons_ad	Adición de consoante	798
O_ac_ad	Adición de acento gráfico	672
M_v_su	Selección dunha forma flexiva verbal non estándar	520

Así pois, as desviacións máis recorrentes son as que afectan ao emprego dos signos de puntuación, os ortográficos relativos á acentuación (*vid.* o capítulo de López Sánchez e Lorenzo Herrera neste volume), á selección do léxico (sobre todo pola presenza de moitas transferencias do español, *vid.* o capítulo de Álvarez de la Granja nesta obra) e ao seu emprego contextual adecuado, á selección do número (esencialmente por concordancias inadecuadas, *vid.* o capítulo de Cidrás neste volume) e á flexión verbal (normalmente, tamén por influencia do español, *vid.* Álvarez de la Granja 2020b). Ademais, existe un considerable número de formas en que se produce a adición dunha consoante (O_cons_ad), pero debe sinalarse que un 85% das palabras etiquetadas con este código son as voces *producto*, *producción* e outras formas da

mesma familia de palabras (*productor, productividade...*), moi recorrentes no corpus dado que un dos comentarios trata sobre o consumo e a produción.

Un segundo nivel de anotación atende á orixe da desviación do estándar, pero como xa indicamos en §2.5.4, o emprego destes códigos está restrinxido a uns poucos casos predeterminados, de xeito que non todas as formas non estándar reciben unha etiqueta deste tipo. Ademais, unha mesma desviación pode ser explicada con máis dun código. En consecuencia, calquera análise e comparativa que se realice está condicionada por estas características. Na Táboa 10¹⁴ ofrécese información sobre a distribución dos códigos asignados, que son un total de 7995¹⁵ e se presentan ordenados de maior a menor frecuencia.

TÁBOA 10. Distribución dos códigos de orixe

Orixe	Códigos	Número de códigos asignados	Porcentaxe
<i>sp</i>	O_sp M_sp G_sp L_sp S_sp	5033	62,95%
<i>gal</i>	M_gal L_gal G_gal S_gal	820	10,26%
<i>spadapt</i>	M_spadapt L_spadapt	649	8,12%
<i>anal</i>	O_anal M_anal L_anal	592	7,40%
<i>hc</i>	L_hc G_hc	576	7,20%
<i>or</i>	G_or	206	2,58%
<i>for</i>	O_for M_for L_for S_for	109	1,36%
<i>mix</i>	L_mix	10	0,13%
		7995	

¹⁴ Na Táboa 4 poden verse os valores e os casos en que se emprega cada elemento.

¹⁵ Tamén desta volta computamos manualmente aqueles casos en que unha mesma etiqueta aparece por duplicado asociada a un *token*, como ocorre, por exemplo, cun caso de colocación enclítica en vez de proclítica dun pronome átono e ao mesmo tempo de cheísmo. Neste caso, o clítico etiquétase con dous códigos G_hc.

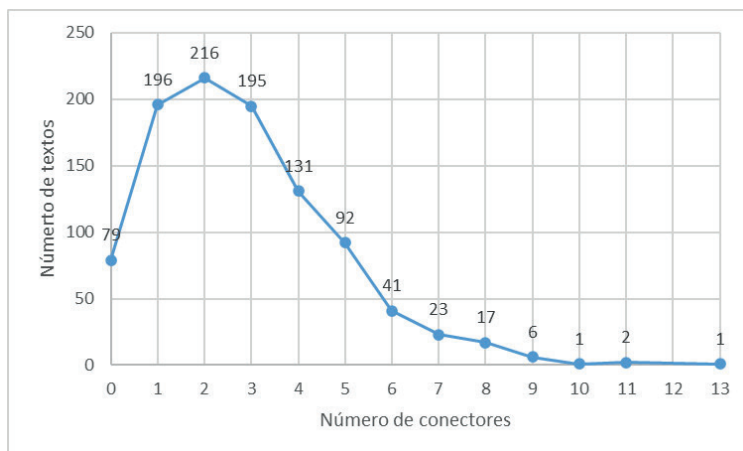
Tal e como se pode comprobar na Táboa 10, as etiquetas máis recorrentes, con moita diferenza sobre as demais, son as que atribúen a orixe da forma non estándar a unha transferencia do español. Precisamente, o código máis frecuente é L_sp, do que hai 2270 exemplos, e que identifica as unidades léxicas transferidas directamente desde o castelán. A segunda orixe máis frecuente é aquela que identifica as formas procedentes de variedades non estándar do galego, sendo M_gal o código máis repetido neste grupo, con 621 resultados. A razón estriba sobre todo na frecuencia no corpus de formas verbais que se poden explicar por transferencia desde o galego popular e tamén pola presenza de numerosos exemplos do demostrativo neutro con <e> (*esto, eso* e en menor medida *aquelo*). Nótese, en calquera caso, que estas últimas formas e unha boa parte das primeiras tamén se poden explicar desde o español e reciben igualmente os correspondentes códigos de orixe. Superan tamén as 500 ocorrencias os casos en que se produce a adaptación dunha unidade léxica ou dun elemento flexivo desde o español, así como as formas que se poden explicar como hipercorreccións e como analoxías.

3.4. Os conectores que vinculan enunciados

Finalmente, ofrecemos os datos cuantitativos sobre os conectores empregados no corpus para vincular enunciados. Son 2839 os códigos deste tipo asignados (o que supón unha media de 2,84 conectores deste tipo por texto). A media en xuño é de 2,93 e en setembro de 2,04, cunha diferenza de case un conector entre unha e outra convocatoria.

O rango sitúase entre 13 conectores nunha redacción (da oportunidade de xuño) e ningún en 79 textos (o 16,67% dos textos de setembro non teñen ningún conector entre enunciados e o 6,90% dos de xuño), mentres que a desviación estándar é de 1,93. Na Gráfica 7 obsérvase a distribución dos conectores entre os 1000 textos. Nela, podemos comprobar como a cantidade máis repetida é 2 (216 textos teñen dous conectores deste tipo) e, a medida que se incrementa a cifra de conectores, diminúe o número de textos.

GRÁFICA 7. Distribución dos conectores nos textos de CORTEGAL



A distribución por tipo é a que se recolle na Táboa 11¹⁶, onde a ordenación atende á frecuencia de cada tipo de conector.

TÁBOA 11. Distribución dos códigos dos conectores que vinculan enunciados

Código	Número de códigos	Porcentaxe
ad	913	32,16%
contraarg	616	21,70%
cons	389	13,70%
concl	356	12,54%
ord_sp	225	7,93%
exempl	100	3,52%
ord_t	98	3,45%
caus	57	2,01%
reform	33	1,16%
coment	29	1,02%
confirm	13	0,46%
disjunc	7	0,25%
fin	3	0,11%
2839		

¹⁶ Na Táboa 5 pode atoparse o valor de cada código.

Tal e como se pode comprobar na Táboa 11, os conectores máis frecuentes son os de adición, seguidos dos contraargumentativos, dos consecutivos e dos conclusivos. Non resultan estraños estes resultados, tendo en conta o carácter argumentativo dos textos. O conector máis frecuente é *tamén*, etiquetado en 421 casos. Os outros conectores que superan os 100 códigos son *pero*, *en conclusión*, *e* e *ademais*. Lémbrese en todo caso, que estes elementos só se etiquetan cando serven para unir enunciados.

4. CONCLUSIÓN

Na primeira parte deste traballo presentamos os criterios metodolóxicos empregados tanto na selección da mostra dos textos que conforman *CORTEGAL* como na transcripción e na anotación das redaccións. Por razóns de espazo, a presentación sobre estas dúas últimas tarefas límitase a aspectos xerais, pero na páxina do corpus poden atoparse diferentes documentos que complementan a información proporcionada aquí.

Ao longo desta primeira parte, tamén demos conta das posibilidades de visualización, tanto dos textos como das súas anotacións, que ofrece *TEITOK*, a plataforma en que se desenvolve o corpus. O feito de poder visualizar os textos nas súas diferentes capas de estandarización, así como a identificación mediante distintas cores das formas non estándar, contribúen a que o corpus poida ser empregado nas aulas para traballar diferentes aspectos lingüísticos.

A anotación das formas non estándar realizada no corpus sitúase no marco metodolóxico da denominada «Análise informatizada de erros» (ou «Análise de erros asistida por ordenador»), pero o deseño do sistema de anotación, para o que non existe un modelo estandarizado, é específico de *CORTEGAL* e responde aos obxectivos do corpus e ás especificidades dos textos que o conforman. Optamos por unha análise bastante detallada das formas non estándar, cun sistema multicapa organizado en seis niveis lingüísticos e xerarquizado desde os niveis onde se levan a cabo estandarizacións vinculadas á forma dos *tokens* ata aqueloutros en que é preciso atender a aspectos contextuais para realizar a normalización. Ademais de codificar o tipo de desviación do estándar (nivel e aspecto afectado, e diferenza entre a forma estándar e a escrita polo alumnado) consideramos de interese informar nalgúns casos, e nun campo de anotación distinto, da orixe das formas diverxentes, así como ofrecer unha primeira aproximación á anotación doutros aspectos discursivos, máis alá das desviacións do estándar, mediante a identificación e codificación dos conectores entre enunciados.

Na segunda parte do traballo presentamos algúns dos resultados atopados en *CORTEGAL*. Limitámonos a aspectos xerais, posto que a análise detallada das distintas cuestións se leva a cabo nos restantes traballos desta mesma obra, en traballos previos (como Álvarez de la Granja, 2020a e Álvarez de la Granja, 2020b) así como, agardamos, en futuras publicacións.

De acordo coa análise presentada, un texto que atenda ás medias de *CORTEGAL* ten ao redor de 225 palabras (113 diferentes), 9 enunciados de 27 palabras cada un, e entre 4 e 5 parágrafos con 2 ou 3 enunciados. Para unir os diferentes enunciados, emprega tan só 3 conectores. Por outro lado, o texto ten 22 ou 23 formas ou secuencias non estándar, unha cifra que parece bastante elevada tendo en conta a extensión das redaccións. En calquera caso, a presentación realizada responde sobre todo ás características dos textos da convocatoria de xuño, moito máis próximos á media dado que supoñen un 89,8% do total das redaccións do corpus. Os textos de setembro conteñen menos palabras, menos lemas ou palabras diferentes e menos enunciados, pero os que inclúen son máis longos e están vinculados con menos conectores. Ademais, un texto desta convocatoria ten 4 formas non estándar máis ca un texto da convocatoria de xuño. Estas diferenzas van ligadas ás que atopamos nas cualificacións, posto que os textos da primeira convocatoria teñen unha nota media considerablemente máis alta ca os da segunda, cunha diferenza que supera os 1,7 puntos.

As desviacións do estándar son sobre todo relativas ao emprego dos signos de puntuación, á acentuación e á selección de léxico (léxico non normativo ou inadecuado ao contexto). Hai tamén algúns problemas relacionados coa concordancia de número, coa ortografía e coa conxugación verbal. Unha boa parte destas desviacións do estándar responden á transferencia ou influencia do castelán. Precisamente, o capítulo de López-Sández e Lorenzo-Herrera deste traballo, que aborda a acentuación, e o capítulo de Álvarez de la Granja, que trata o léxico non estándar, afondan sobre esta cuestión e amosan ben ás claras a relevancia cuantitativa da influencia do español nas desviacións da norma.

Non queremos rematar este traballo sen poñer de manifesto que a complexa anotación levada a cabo en *CORTEGAL* é susceptible, evidentemente, de correccións e ampliacións. Por un lado, tal e como indicamos en §2.5.6, non se nos escapa a existencia de posibles erros e incoherencias na anotación realizada, eivas e inconsistencias que iremos corrixindo a medida que sexan detectadas. Por tal motivo, as cifras que ofrecemos aquí poden non coincidir exactamente cos resultados que proporcione o corpus en futuras buscas,

aínda que, en calquera caso, os cambios non serán significativos. Por outro lado, consideramos de interese ampliar a anotación e refinala, cando cumpra, en futuras versións do corpus, tanto no relativo ás formas non estándar como a outros aspectos que nos axuden a coñecer as características da escrita en lingua galega do estudantado de Galicia ao remate da educación secundaria, particularmente en relación cos aspectos discursivos.

Tanto o corpus, co seu sistema de buscas, como os documentos empregados ao longo da súa elaboración están dispoñibles en aberto para calquera persoa, sexa persoal investigador, profesorado de lingua galega ou calquera outra/o usuaria/o interesada/o. Así mesmo, as persoas responsables do corpus estamos abertas ás súas indicacións e suxestións tanto para correccións de erros como para melloras e ampliacións de máis longo alcance.

O noso obxectivo cando comezamos coa elaboración de *CORTEGAL* en 2017 era deseñar unha ferramenta que servise para coñecer, con datos obxectivos e representativos, como escribe o noso alumnado e, particularmente, aínda que non só, para detectar as principais eivas e carencias na destreza da escritura en lingua galega estándar, pois partiamos da hipótese, demostrada en traballos previos (Silva Valdivia, 2010; López Meirama e Álvarez de la Granja, 2014) e corroborada posteriormente (Loredó e Silva Valdivia, 2020), de que existían algunhas deficiencias importantes en distintos niveis lingüísticos. A análise dos datos confirma unha vez máis esta hipótese, pero *CORTEGAL*, ademais, permite manexar de primeira man os textos que sustentan a súa veracidade. Agardamos que este recurso contribúa tanto a tomar conciencia da necesidade de adoptar medidas para mellorar a destreza da escritura na variedade estándar da lingua galega como a participar directamente nesa mellora mediante a súa utilización directa nas aulas.

REFERENCIAS BIBLIOGRÁFICAS

- ABEL, A., GLAZNIEKS, A. e CULY, C. (2012). *KoKo German L1 Learner Corpus v2*, Eurac Research CLARIN Centre. <http://hdl.handle.net/20.500.12124/11>
- ABEL, A., GLAZNIEKS, A., NICOLAS, L. e STEMLE, E. (2014). KoKo: an L1 Learner Corpus for German. En N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk e S. Piperidis (Eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)* (pp. 2414-2421). European Languages Resources Association.
- ABEL, A., GLAZNIEKS, A., NICOLAS, L. e STEMLE, E. (2016). An extended version of the KoKo German L1 Learner corpus. En A. Corazza, S. Montemagni e G. Semeraro (Dirs.), *Proceedings of the Third Italian Conference on*

- Computational Linguistics CLiC-it 2016* (pp. 13-18). Accademia University Press. <https://doi.org/10.4000/books.aaccademia.1743>
- ALONSO-RAMOS, M. (2016). Spanish learner corpus research. Achievements and challenges. En M. Alonso-Ramos (Ed.), *Spanish Learner Corpus Research Current trends and future perspectives* (pp. 3-31). John Benjamins. <https://doi.org/10.1075/scl.78.01alo>
- ÁLVAREZ DE LA GRANJA, M. (2018). Corpus de textos de estudantes galegos (CORTEGAL). Aspectos metodolóxicos. En M. Díaz, G. Vaamonde, A. Varela, M. C. Cabeza, J. M. García-Miguel e F. Ramallo (Eds.), *Actas do XIII Congreso Internacional de Lingüística Xeral* (pp. 55-62). Universidade de Vigo. <http://cilx2018.uvigo.gal/actas/resumos/655842.html>
- ÁLVAREZ DE LA GRANJA, M. (2020a). Análise da conxugación verbal no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL). En L. de Castro Moutinho, R. L. Coimbra e A. Gómez Bautista (Coords.), *Línguas Minoritárias e Variação Lingüística* (pp. 150-173). Universidade de Aveiro. <https://doi.org/10.34624/rj68-vz44>
- ÁLVAREZ DE LA GRANJA, M. (2020b). O infinitivo flexionado no *Corpus de textos galegos escritos por estudantes no ámbito académico* (CORTEGAL). *Madrygal. Revista de estudos gallegos*, 23, 15-38. <https://doi.org/10.5209/madr.73064>
- ÁLVAREZ DE LA GRANJA, M. e LÓPEZ MEIRAMA, B. (2021). La presencia del español en el léxico disponible del gallego. El centro de interés el cuerpo humano. En M. Serrano Zapata e M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 115-145). Tirant Humanidades.
- AMARO, R., CORREIA, S., GRAMACHO, C. e MENDES, A. (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>
- BURNARD, L. (2005). Metadata for corpus work. En M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice*. Oxbow Books. <https://users.ox.ac.uk/~martinw/dlc/chapter3.htm>
- CAPSADA BLANCH, R. e TORRUELLA CASAÑAS, J. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba*, 44, 47-408. <https://doi.org/10.15304/verba.44.3155>
- CENTRE FOR ENGLISH CORPUS LINGUISTICS (2022). *Learner Corpora around the World*. Université catholique de Louvain. <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- CORDER, S. P. (1974). Error analysis. En J. P. Allen e S. P. Corder (Eds.), *The Edinburgh Course in Applied Linguistics* (pp. 122-154, vol. 3). Oxford University Press.
- DAGNEAUX, E., DENNESS, S. e GRANGER, S. (1998). Computer-aided error analysis. *System*, 26, 126-174. [https://doi.org/10.1016/S0346-251X\(98\)00001-3](https://doi.org/10.1016/S0346-251X(98)00001-3)

- DÍAZ-NEGRILLO, A. e FERNÁNDEZ DOMÍNGUEZ, J. (2006). Error Tagging Systems for Learner Corpora. *Revista española de lingüística aplicada*, 19, 83-102.
- DÍEZ-BEDMAR, M. B. (2021). Error Analysis. En N. Tracy-Ventura e M. Paquot (Eds.), *The Routledge Handbook of Second Language Acquisition and Corpora* (pp. 99-104). Routledge. <https://doi.org/10.4324/9781351137904-9>
- DULAY, H., BURT, M. e KRASHEN, S. (1982). *Language two*. Oxford University Press.
- ELLIS, R. (1994). *The Study of Second Language Acquisition*. Oxford University Press.
- FIGUERAS, C. (1999). La semántica procedimental de la puntuación. *Espéculo. Revista de estudios literarios*, 12. <https://webs.ucm.es/info/especulo/numero12/puntuac.html>
- GILQUIN, G. e GRANGER, S. (2022). Using data-driven learning in language teaching. En A. O’Keeffe e M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (2.^a ed., pp. 430-442). Routledge. <https://doi.org/10.4324/9780367076399>
- GÓMEZ GUINOVART, X. e SOLLA, M. (2017-2022). *DContado*. Universidade de Vigo. <https://ilg.usc.gal/dcontado/index.php?lang=gl>
- GONZÁLEZ ÁLVAREZ, E. (1999). Análisis de los errores léxico-semánticos. En L. 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* (pp. 207-270). Universidade de Santiago de Compostela.
- GRANGER, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20(3), 465-480. <https://doi.org/10.1558/cj.v20i3.465-480>
- GRANGER, S. (2017). Learner Corpora in Foreign Language Education. En S. Thorne e S. May (Eds.), *Language, Education and Technology. Encyclopedia of Language and Education* (pp. 427-440). Springer. https://doi.org/10.1007/978-3-319-02237-6_33
- GRANGER, S., GILQUIN, G. e MEUNIER, F. (2015). Introduction: Learner corpus research - Past, present and future. En S. Granger, G. Gilquin e F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 1-5). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.001>
- INSTITUTO GALEGO DE ESTATÍSTICA (2019). *Enquisa estrutural a fogares. Coñecemento e uso do galego*. Instituto Galego de Estatística. https://www.ige.eu/web/mostrar_actividade_estadistica.jsp?idioma=gl&codigo=0206004
- JANSSEN, M. (2016). TEITOK: Text-Faithful Annotated Corpora. En N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk e S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 4037-4043). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2016/pdf/651_Paper.pdf

- LEECH, G. e WILSON, A. (1996). *Recommendations for the Morphosyntactic Annotation of Corpora*. Expert Advisory Group on Language Engineering Standards. <https://home.uni-leipzig.de/burr/Verb/htm/LinkedDocuments/annotate.pdf>
- Lei 3/1983 de 15 de xuño de Normalización lingüística. 14 de xullo de 1983. *Diario Oficial de Galicia*, nº 84.
- LÓPEZ MEIRAMA, B. e ÁLVAREZ DE LA GRANJA, M. (2014). *Léxico dispoñible do galego*. Universidade de Santiago de Compostela. <https://www.usc.gal/libros/es/lingstica/125-lexico-disponible-do-galego-lexico-disponible-do-galego.html>
- LOREDO, X. e SILVA VALDIVIA, B. (Coords.) (2020). *Avaliación da competencia bilingüe nos idiomas galego e castelán do alumnado de 4º da ESO*. Real Academia Galega. <https://publicacions.academia.gal/index.php/rag/catalog/view/372/366/383>
- LÜDELING, A., ADOLPHS, P., KROYMANN, E. e WALTER, M. (2005). Multi-level error annotation in learner corpora, *Proceedings from the Corpus Linguistics Conference Series* 1(1). University of Birmingham. <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>
- LÜDELING, A. e HIRSCHMANN, H. (2015). Error annotation systems. En S. Granger, G. Gilquin e F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 135-158). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.007>
- MCENERY, T. e RICHARD, X. (2010) What corpora can offer in language teaching and learning. En E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 364-380, vol. 2). Routledge.
- MERLIN project (2014). *Annotation guidelines*. <https://merlin-platform.eu/docs/Annotation%20guidelines.pdf>
- MENDES, A., ANTUNES, S., JANSSEN, M. e GONÇALVES, A. (2016). The COPLE2 corpus: A learner corpus for Portuguese. En N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odiijk e S. Piperidis (Eds.), *Proceedings of the Tenth Language Resources and Evaluation Conference (LREC 2016)* (pp. 3207-3214). European Language Resources Association (ELRA). <https://aclanthology.org/L16-1511.pdf>
- MIKELIĆ PRERADOVIĆ, N. (2020). Označavanje pogrešaka u CroLTeC-u (računalnom učeničkom korpusu hrvatskog kao stranog jezika). *Rasprave: Časopis Instituta za Hrvatski Jezik i Jezikoslovlje*, 46(2), 899-920. <https://doi.org/10.31724/rihjj.46.2.24>
- REZNICEK, M., LÜDELING, A., KRUMMES, C., SCHWANTUSCHKE, F., WALTER, M., SCHMIDT, K., HIRSCHMANN, H. e ANDREAS, T. (2012). *Das Falko-Handbuch: Korpusaufbau und Annotationen. Version 2.01*. Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/view>

- REZNICEK, M., LÜDELING, A. e HIRSCHMANN, H. (2013). Competing target hypotheses in the Falko corpus. A flexible multi-layer corpus architecture. En A. Díaz-Negrillo, N. Ballier e P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 101-124). John Benjamins. <https://doi.org/10.1075/scl.59.07rez>
- RÍO, I. DEL e MENDES, A. (2018). Error annotation in the COPLE2 corpus. *Revista da Associação Portuguesa de Linguística*, 4, 225-239. <https://doi.org/10.26334/2183-9077/rapln4ano2018a42>
- RÖMER, U. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics*, 31, 205-225. <https://doi.org/10.1017/S0267190511000055>
- ROSELLÓ VERDEGUER, J. (2015). El texto y sus propiedades: algunas consideraciones de carácter práctico. *Tonos Digital. Revista de estudios filológicos*, 28. https://www.um.es/tonosdigital/znum28/secciones/tintero-7--rosello_texto.htm
- ROSEN, A. (2015). *CzeSL-MAN - a corpus of non-native speakers' Czech with manual annotation*. Informe técnico. Charles University in Prague. <http://utkl.ff.cuni.cz/~rosen/public/2015-czesl-man-en.pdf>
- STEMLE, E. W., BOYD, A., JANSSEN, M., LINDSTRÖM TIEDEMANN, T., MIKELIĆ PRERADOVIĆ, N., ROSEN, A., ROSÉN, D. e VOLODINA, E. (2019). Working together towards an ideal infrastructure for language learner corpora. En A. Abel, A. Glaznieks, V. Lyding e L. Nicolas (Eds.), *Widening the Scope of Learner Corpus Research. Selected papers from the fourth Learner Corpus Research Conference* (pp. 427-468). Presses universitaires de Louvain.
- SILVA VALDIVIA, B. (Dir.). (2010) *Avaliación da competencia do alumnado de 4º da ESO nos idiomas galego e castelán*. Universidade de Santiago de Compostela.
- TENFJORD, K., MEURE, P. e HOFLAND, K. (2006). The ASK corpus: A language learner corpus of Norwegian as a second language. En N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk e D. Tapias (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1821-1824). European Languages Resources Association. http://www.lrec-conf.org/proceedings/lrec2006/pdf/573_pdf.pdf
- TIMMIS, I. (2015). *Corpus Linguistics for ELT: Research and Practice*. Routledge. <https://doi.org/10.4324/9781315715537>
- VÁZQUEZ ROZAS, V. e BLANCO, M. (2022). Corpus y enseñanza del español. En G. Parodi, P. Cantos-Gómez e C. Howe (Eds.), *Linguística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 342-356). Routledge. <https://doi.org/10.4324/9780429329296-26>

Análise da acentuación gráfica a través do corpus *CORTEGAL*

MARÍA LÓPEZ-SÁNDEZ
Universidade de Santiago de Compostela

LARA LORENZO-HERRERA
Universidade de Santiago de Compostela

1. INTRODUCCIÓN

A acentuación gráfica constitúe un dos aspectos que concentra un número máis elevado de anotacións en *CORTEGAL* (vid. o capítulo de Álvarez de la Granja e Rodríguez Rodríguez neste volume, §3.3), o que responde á alta prevalencia dos erros de acentuación no conxunto das desviacións da norma do alumnado galego. A análise a través do corpus *CORTEGAL* permite detectar cales son os puntos do sistema nos que se concentran as dificultades do alumnado e extraer conclusións relevantes para o deseño da intervención orientada á súa corrección.

Na xerarquía de gravidade entre os erros ortográficos, adoita conferirse maior importancia á ortografía regrada ca á arbitraria e a aquela que implica cambios con valor de discriminación distintiva sobre a que non. Segundo ambos os parámetros, os erros de acentuación entran nunha categoría elevada (Figuroa de Amorós, 1994, p. 91; González Pascual, 2003, p. 489), superior á da ortografía arbitraria que pode afectar a <b/v>, uso indebido de <h> ou outros erros semellantes, posto que a acentuación ten implicacións fonemáticas e distintivas e está regulada por unha norma consistente e sistemática. Así, segundo unha tipoloxía clásica do erro ortográfico (Polo, 1974), a acentuación ocuparía o segundo lugar en gravidade, directamente despois dos erros de puntuación que afectan á estrutura sintáctica e textual.

Esta importancia outorgada por estudiosos e lingüistas non coincide coa percepción maioritaria da poboación, que tende a sancionar máis duramente ou reaccionar con maior forza fronte a cambios na ortografía arbitraria como os que afectan a <b/v> ou <h>. Isto mesmo observa González Pascual (2003) respecto aos usos xornalísticos da linguaxe nos que centra o seu estudo dos erros da acentuación, pois detecta unha preocupación en evitar nos xornais os outros tipos de erro mais un desleixo considerable no tocante á acentuación. Na propia percepción espontánea do erro de acentuación, especialmente o consistente na omisión, inflúe sen dúbida a proliferación actual de formas de escrita dixital marcadas pola rapidez e inmediatez, tendente á omisión dunha marca ortográfica que non resulta áxil de inserir en moitos dispositivos. Malia todo isto, a sistematicidade da norma que a regula e a súa implicación fonemática apoian o interese por deseñar estratexias didácticas e metodolóxicas que minimicen os erros e reduzan as dificultades do alumnado. Resulta pois esencial coñecer as tendencias e os puntos do sistema nos que se concentran os desvíos, para o que os corpus de textos de estudantes como *CORTEGAL* constitúen unha ferramenta de máxima relevancia.

2. ANÁLISE DOS FENÓMENOS DE OMISIÓN, ADICIÓN E DESPRAZAMENTO

Para a análise dos datos a través do corpus establecéronse tres categorías xerais no que atinxe á colocación do acento gráfico: casos de omisión, adición e desprazamento, que abranguen todas as posibilidades de erro na aplicación do acento gráfico e que contan con precedentes na bibliografía sobre o erro ortográfico e a acentuación (Álvarez, 1977; Tomcsányi, 2015; Sotomayor et al., 2017). Déixanse fóra da análise aqueles casos nos que hai razóns para pensar que non se trata dun problema de aplicación da norma acentual, senón dun cambio no acento de intensidade, como acontece en termos cunha tonicidade diferente á do castelán (comentados no capítulo de Álvarez de la Granja, neste volume, §3.2.3) ou nos casos de acentuación proparoxítona dos verbos en copretérito ou pospretérito. Así pois, cómpre precisar que neste traballo se considera erro de acentuación calquera desvío da norma imputable á aplicación das regras de acentuación. A omisión (1) implica a supresión do acento gráfico na vogal tónica da sílaba correspondente. A adición (2) supón a incorporación do acento na vogal tónica dunha palabra á que, segundo as normas de acentuación, non lle corresponde levalo e, por último, o desprazamento consiste no cambio de posición do acento gráfico nunha palabra, de forma que este é colocado sobre unha vogal átona (3). A continuación, amósanse exemplos dos tres tipos de erros pertencentes á mostra de *CORTEGAL*:

- (1) Por outro lado, tamén cabe destacar que hai moita xente de paladar plano e prefire non arriesgar con sabores e texturas **deixándose** levar polo **clasico** ovo frito
- (2) As cousas que intentanos vender como a creatividade, a fusión dos sabores, quizais queden no olvido cando a xente volva a lembrar o rico que están os pratos da súa **avó**a ou a posibilidade de comer até fartar nas tabernas en vez de esperar anos para comer nun restaurante e pagando grandes prezos
- (3) Por exemplo cada vez hai **maís** programas e concursos de cociña, tanto para adultos como para pequenos e de case todas as partes do mundo

A distribución dos erros de colocación evidencia un claro dominio dos erros por omisión (76,51%), mentres que a adición dun acento gráfico indebido implica algo máis dun de cada cinco erros rexistrados (21,55%), sendo a dislocación acentual case residual ao representar menos de dous de cada 100 erros (1,94%). O feito de que a adición ou o desprazamento dun acento teña un efecto máis rechamante que a mera omisión propicia que a estratexia maioritaria ante a dúbida sexa a omisión. Pola súa parte, o desprazamento do acento gráfico é o erro que compromete máis gravemente a acentuación, pois para que se produza debe haber unha dificultade na percepción da tonicidade, de xeito que a aplicación do acento se realiza de modo mecánico e non funcional. A adición, pola súa parte, constitúe unha forma de hipercorrección ortográfica que manifesta, ademais do problema co manexo da acentuación, a conciencia desa dificultade e un intento de superala ou compensala que acaba implicando erros por ultracorrección. Con todo, isto sería así de xeito estrito nun contexto no que só fose operativo un sistema de acentuación, mais a situación complícase cando a convivencia de dúas linguas ambientais, con todas as implicacións sociolingüísticas e didácticas que afectan ao galego e castelán en Galicia, introduce o factor de interferencia entre dúas normas parcialmente coincidentes mais tamén diverxentes en puntos importantes do sistema. Así, a alta prevalencia do fenómeno de adición do acento atopa a explicación, como veremos, na interferencia do sistema acentual do castelán.

O número de erros por desprazamento, de acordo co indicado, é baixo, con 60 ocorrencias de formas como **límitanse*, **futból*, **maís*, **acabáanse*, **dasé*, **existírian*, **póderia* ou **saúdabel*. Fronte a estes 60 casos, rexístranse 2360 de omisión. Con todo, para realizar a análise pormenorizada e cualitativa dos erros son sen dúbida os casos de adición (665) os máis relevantes, pois esa adición non é, como tende a acontecer nos limitados casos de desprazamento, resultado da dificultade coa percepción da tonicidade ou mesmo froito dun simple lapsus cáلامي, senón que derivan na súa maioría dunha aplicación errada da norma, a miúdo como resultado da interferencia co

sistema acentual do castelán. Rexístranse en *CORTEGAL* erros coma *avóa, *quére, *éla, *artistas, *cadéa, *estár, *ántes, *séus, *inváden, *mantér, *moréas ou *búscan, nos que se produce a adición dun acento gráfico sobre a vogal tónica contravindo as regras xerais de acentuación sen que a norma do castelán explique a adición, pero estes exemplos son residuais fronte aos que se poden explicar por esa interferencia.

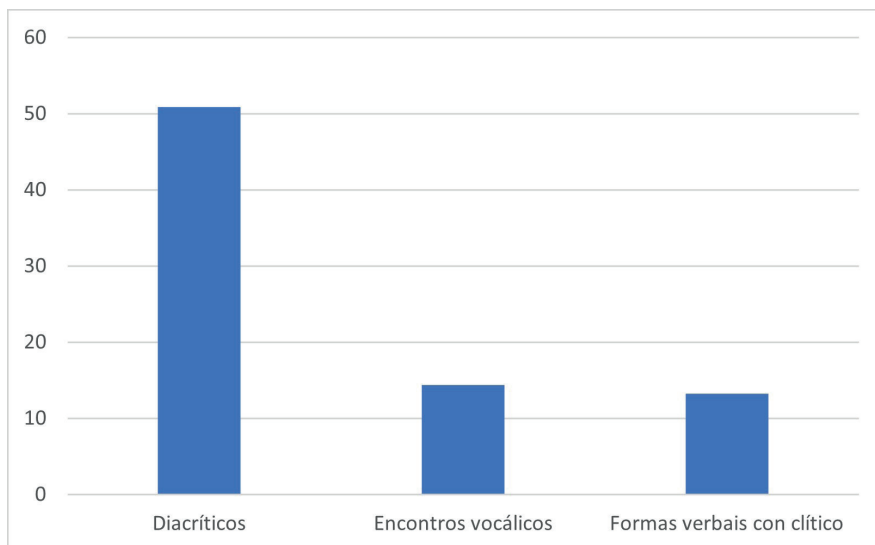
A conxugación dos factores cuantitativo e cualitativo na análise dos erros apunta a que a maioría dos desvíos por adición son imputables a interferencias entre os sistemas de acentuación do galego e o castelán. Afondando na análise a través das omisións, isto vese corroborado pola alta prevalencia da omisión do acento diacrítico así como o elevado número de formas verbais con clítico que requirirían acento pola estrutura acentual resultante e nas que se omite o acento gráfico. Para unha análise máis pormenorizada dos erros de omisión, tivéronse en conta aqueles que resultan máis facilmente clasificables, ben polo tipo de palabra ao que pertencen ou ben porque se atopan nalgũa das tipoloxías referidas por Solla Portela (2011) respecto ás diferenzas de acentuación entre o castelán e o galego¹. As categorías escollidas foron os diacríticos, as formas verbais con clítico e os encontros vocálicos.

A análise cuantitativa destes fenómenos ratifica o peso da interferencia no volume total de desvíos da norma. Así, os diacríticos representan a maioría dos erros de acentuación tanto na omisión coma nos outros fenómenos de adición e desprazamento. As formas verbais con clítico constitúen unha categoría diferencial do galego respecto ao castelán na que se rexistran, en xeral, grandes dificultades para o alumnado. Para esta análise cuantitativa, incluíronse para os encontros vocálicos todos os casos de confluencia entre vogais abertas e pechadas e todos os casos de «iu-ui», tanto os que conforman hiatos coma ditongos.

Computáronse un total de 1854 formas clasificables nestas categorías, o que representa o 78,53% dos erros de omisión. Na Gráfica 1 amósanse os resultados, onde o 50,87% dos erros se corresponde coa acentuación dos diacríticos, mentres que as formas nas que existen encontros vocálicos e as formas verbais con clíticos supoñen un 14,40% e un 13,26%, respectivamente.

¹ Solla Portela (2011, p. 59) sintetiza da seguinte maneira as diferenzas de acentuación entre galego e castelán, no seu caso coa finalidade de afinar o software da aplicación Freeling para unha correcta acentuación no galego: «[A]s regras de acentuación gráfica do español difiren das do galego na acentuación diacrítica, na silabación de certos encontros vocálicos (español *atribuimos* / galego *atribuímos*, *atribuíu*, *atribuíamos*) e na consideración das secuencias polisilábicas que rematan en ditongo decrecente ou en ditongo decrecente seguido de -n ou -s (español *comeréis*, *fuereis* / galego *comerei*, *amábeis*). Ademais, os encontros cos pronomes enclíticos presentan particularidades propias».

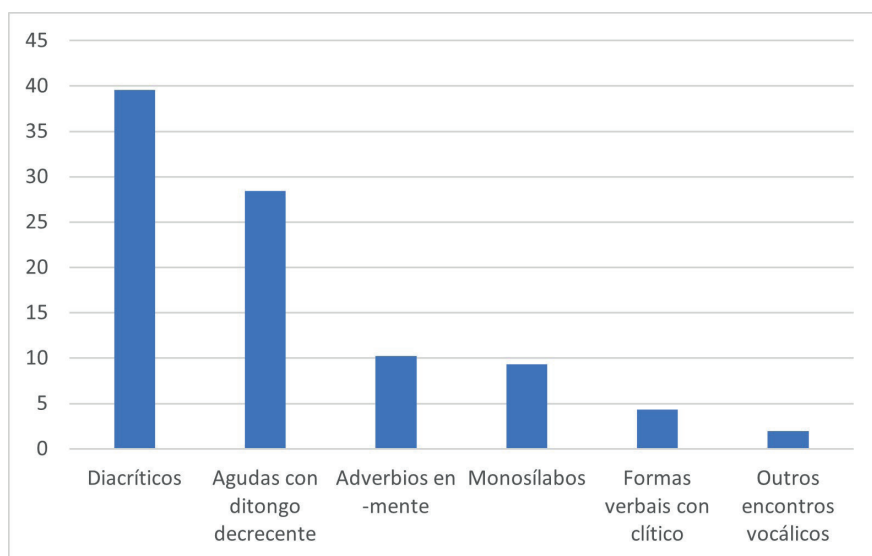
GRÁFICA 1. Clasificación dos erros de omisión



No caso dos diacríticos, rexístrase 453 veces a omisión do acento gráfico sobre as formas contractas da preposición *a* cos artigos, 279 sobre a forma verbal *é*, 166 omisións do acento gráfico sobre o adverbio de cantidade *máis*, 73 sobre a forma verbal *dá*, 69 sobre o pronome tónico *nós*, 60 sobre a forma do verbo *vir vén*, 58 sobre o adxectivo *só* ou 20 sobre o adverbio de lugar *fóra*. Nas formas verbais con clítico, as máis numerosas resultan **basease* (9 ocorrencias), **podese* (8 ocorrencias), **gustalle* (6 ocorrencias) e as formas **gustanos*, **pareceme*, **aproveitanse*, **convertese*, **tratase* e **valorase* con 4 ocorrencias en cada caso. En canto aos encontros vocálicos, cabe destacar as formas **ainda* con 43 casos, **sua* con 27, **países* con 21, **gastronomia* con 14 e **dia* con 12.

No que atinxe ás adicións, computouse un 93,83% dos casos (624 formas) e estableceuse a seguinte tipoloxía: diacríticos, palabras agudas con ditongo decrecente, adverbios en *-mente*, monosílabos, formas verbais con clítico e outros encontros vocálicos. Os datos figuran na Gráfica 2, na que predomina a acentuación diacrítica.

GRÁFICA 2. Clasificación dos erros de adición



A análise detallada dos casos permite destacar estas tendencias:

— Acentuación diacrítica: os erros de adición ligados a este tipo de acentuación son os que manifestan unha frecuencia máis elevada, a cal chega ao 39,55%. É esperable que este punto do sistema acentual resulte especialmente sensible ao erro. A adición do acento na forma gramatical á que non lle corresponde explica unha parte significativa dos desvíos (**máis*, **é*, **ás*, **fóra*, **vén*), o que apunta á dificultade de discriminar entre as formas gramaticais diferenciadas mediante a acentuación diacrítica, mais cómpre observar que en boa medida os erros están motivados na súa maioría novamente polo feito de ser este un punto do sistema con diferenzas evidentes co castelán. En 89 ocasións rexístrase a etiqueta correspondente á acentuación incorrecta de **sí*, que obedece sen dúbida á interferencia, pero o mesmo pode dicirse da altísima recorrencia da adición de acentos derivados da interferencia coa acentuación en castelán de interrogativos directos ou indirectos (**cómo*, **qué*, **quén*, **cómo*...) ou pronomes persoais (**él*). É salientable o número de casos nos que se acentúan demostrativos coma **éste*s ou **ésta* que nunca levaron acento gráfico en galego e para os que tampouco no castelán, desde a última reforma, se recomenda a acentuación, malia o cal presentan 14 etiquetas de erro na mostra.

En consecuencia, é preciso subliñar que nesta categoría se contabilizaron tanto os casos de formas diacríticas propias do galego como aqueles casos que se explican pola interferencia co castelán.

— Acentuación de agudas con ditongo decrecente: outra porcentaxe moi significativa de casos de adición de acento (28,42%) responde a unha incorrecta acentuación das palabras agudas que presentan un ditongo decrecente na última sílaba. Unha forma que concentra un elevado número de erros de acentuación é **ademáis*, que pola súa condición de marcador discursivo se emprega con moita frecuencia e aparece a miúdo incorrectamente acentuado, documentándose 124 ocorrencias, se temos en conta a súa aparición con maiúscula a comezo de enunciado e as formas con minúscula. En 19 ocasións rexístrase **quizáis*; en 3, **xamáis*, ás que cómpre engadir casos coma **especiáis*, **animáis* ou **despóis*.

— Adverbios en *-mente*: é altamente prevalente na mostra o erro consistente en acentuar os adverbios rematados en *-mente* de acordo coa norma operativa en castelán, chegando algúns deles (**únicamente*, **económicamente*, **prácticamente*, **últimamente*) a superar as 5 etiquetas e acadando ata as 10 ocorrencias na mostra, o que implica un 10,23% do total de adicións. A listaxe é ampla: **rápidamente*, **gastronómicamente*, **lógicamente*, **paradóxicamente*, **únicamente*, **íntegramente*, **íntimamente*, **hipócritamente*, **básicamente*, **comúnmente*, **dificilmente*, **drásticamente*... Nalgúns casos a forma errada non coincide plenamente coa aplicación da norma castelá, senón que esixe para a súa explicación a aplicación simultánea da norma do galego e do castelán, como ocorre coa forma **gratuitamente*, na que se aplica a norma galega para a acentuación da secuencia vocálica das vogais cerradas, pero a do castelán no relativo á actuación nos adverbios en *-mente*. Máis complexo aínda resulta explicar un caso como **posiblemente*, no que semella conxugarse a influencia da forma galega *posíbel* coa interferencia da acentuación castelá.

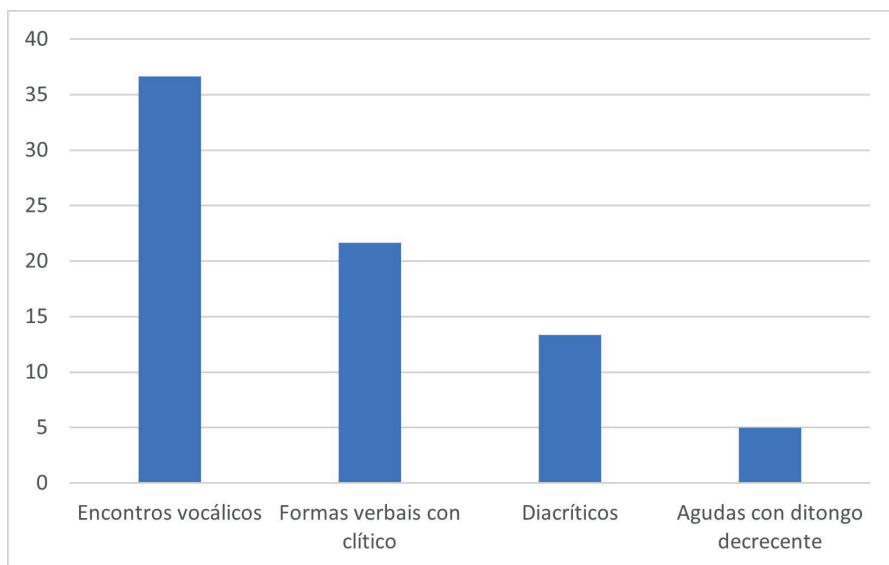
— Monosílabos: un fenómeno observable no corpus que probablemente haxa que poñer en relación coa acentuación diacrítica é a abundancia de adición de acentos sobre formas monosílabas, que supón unha porcentaxe do 9,32%: **bó/s*, **sá*, **més*, **grán*, **crú*, **sán*, **dí*, **máns*, **ún*, **nón*, **dán*, **fái*, **tén*. Se a non acentuación de monosílabos fose sistemática probablemente o número de erros baixaría; mais o feito de que a acentuación diacrítica altere este factor propicia sen dúbida a aparición deste tipo de desvíos da norma.

— Formas verbais con clíticos: as formas verbais con clíticos tamén plasman o fenómeno da adición, aínda que nunha marxe moi inferior ás omisións (4,36%). Rexístranse casos coma **acabóuse*, **afectarános* ou **preguntárlle*, entre outros.

— Outros encontros vocálicos: nesta categoría clasificáronse os casos de encontros vocálicos en palabras que non son agudas rematadas con ditongo decrecente. Malia a súa importancia marxinal (1,95%), mantívose na gráfica para gardar a coherencia na análise respecto ás categorías empregadas nas omisións e nos desprazamentos. Algúns exemplos son **avóa*, **cadéa* ou **persóas*.

En canto aos desprazamentos, contabilizáronse 46, que supoñen un 76,67% do total de erros con esta etiqueta. Aplicáronse catro categorías: formas verbais con clítico, diacríticos, encontros vocálicos e agudas rematadas en ditongo decrecente. Os resultados figuran na Gráfica 3:

GRÁFICA 3. Clasificación dos erros de desprazamento

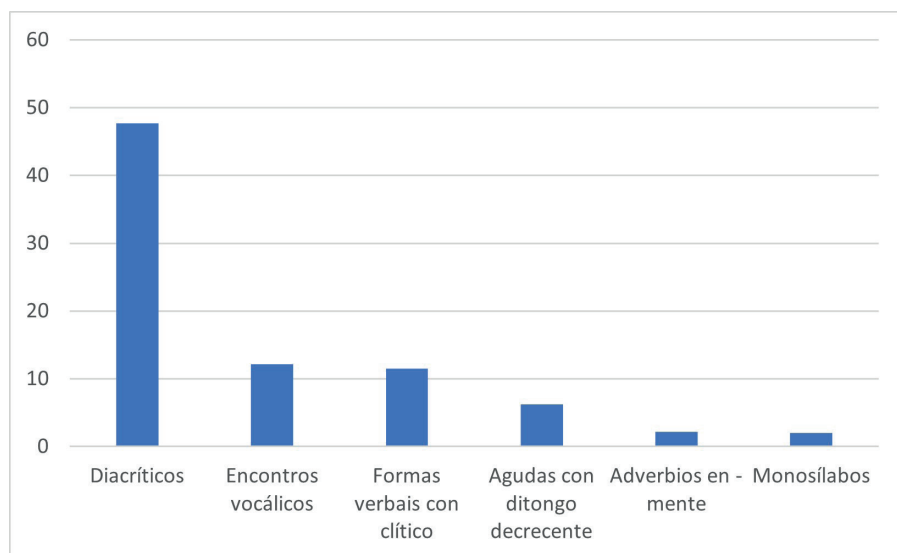


A diferenza do que acontece coas omisións e as adicións, non son os diacríticos os que se atopan en primeiro lugar nos desprazamentos, senón que predominan os encontros vocálicos, os cales acadan unha porcentaxe do 36,67%. Resulta ilustrativo o feito de que na maioría das formas (15 casos) o acento de intensidade cae na vogal que vai ao lado da acentuada graficamente polo estudante (casos como **país*, **saén*, **actúar* ou **doús*), mentres que nas restantes formas (7) non se observa esta tendencia (casos como **continúo*, **industria* ou **saúdabel*). Respecto ás formas verbais con clítico, estas ocupan o segundo lugar, cunha frecuencia do 21,67% e nesta categoría aparecen casos como **centráse*, **entendemonós* ou **podéllo*. Os diacríticos, pola súa

parte, representan o 13,33% dos desprazamentos e todas as ocorrencias se corresponden co cambio na acentuación do adverbio *máis* por **máis* (8 casos). En último lugar, as agudas rematadas en ditongo decrecente presentan unha porcentaxe do 5% e unicamente se rexistran 3 casos, os cales seguen o padrón xa constatado nos encontros vocálicos, o de acentuación gráfica da vogal que se sitúa ao lado da que coincide co acento de intensidade: **ade-máis* (2 ocorrencias) e **demaís*.

Se xuntamos os resultados das Gráficas 1, 2 e 3, obtemos os datos globais da Gráfica 4, nos cales se agruparon os casos analizados dentro de cinco grandes categorías que marcan as diferenzas entre a acentuación en galego e castelán segundo Solla Portela (2011), ás que se engadiron os adverbios en *-mente* e os monosílabos pola súa relevancia: diacríticos, encontros vocálicos, formas verbais con clítico, palabras agudas rematadas en ditongo decrecente, adverbios en *-mente* e monosílabos.

GRÁFICA 4. Clasificación dos erros de acentuación segundo as diferenzas de acentuación entre o galego e o castelán



As frecuencias relativas, desta vez, calculáronse sobre o total de erros de acentuación de *CORTEGAL* (3086), o que mostra un panorama onde sobresaen de xeito claro as dificultades de acentuación cos diacríticos, cunha porcentaxe dun 47,7%. Os encontros vocálicos representan un 12,15% e están seguidos moi de preto polas formas verbais con clítico, cun 11,5%. As agudas

rematadas en ditongo decrecente chegan ao 6,22%, os adverbios en *-mente* manifestan un 2,2% e os monosílabos reflicten un 2,01%. Todos estes datos supoñen o 81,79% da mostra.

Para a análise, considerouse á parte a categoría das agudas con ditongo decrecente porque estas destacan coma un erro frecuente principalmente nas adicións e en moita menor medida nos desprazamentos, mais non deixan de ser unha subclase dos encontros vocálicos. Isto implica que, de terse computado as agudas con ditongo decrecente no mesmo grupo que os encontros vocálicos, estes seguirían ocupando o segundo lugar pero cuns valores máis elevados, do 18,37%.

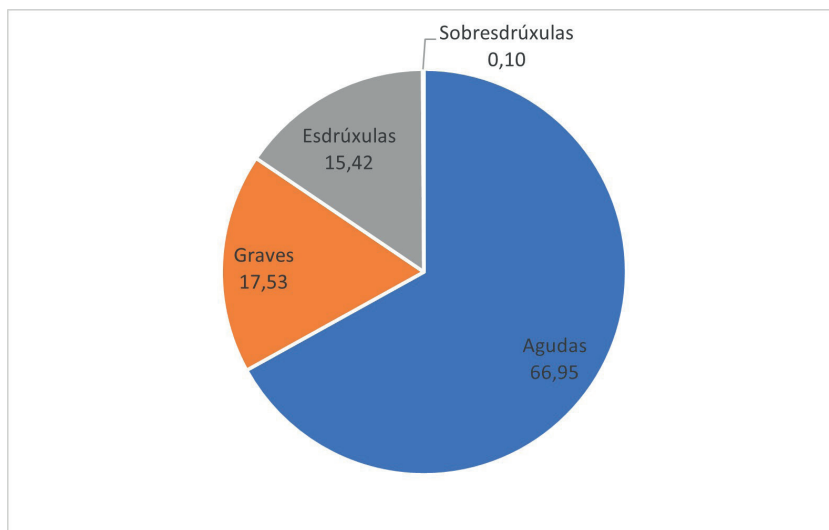
Doutra banda, os adverbios en *-mente* e os monosílabos resultan especialmente significativos en canto ao seu peso cuantitativo, xa que só se rexistran nunha dinámica, nas adicións, pero supoñen un 2,2% e un 2,01% sobre o total de erros de acentuación. Pola contra, os diacríticos, as formas verbais con clítico e os encontros vocálicos encóntranse nos tres fenómenos de colocación do acento examinados.

En consecuencia, o 81,79% dos erros de acentuación analizados teñen algún tipo de relación directa ou indirecta coas interferencias entre os sistemas de acentuación do galego e do castelán. Isto significa que a aplicación das normas de acentuación do galego está fondamente interferida pola norma do castelán e que hai unha escasa conciencia das diferenzas que distinguen os dous sistemas.

3. ANÁLISE DE ERROS SEGUNDO A ESTRUTURA ACENTUAL DA PALABRA

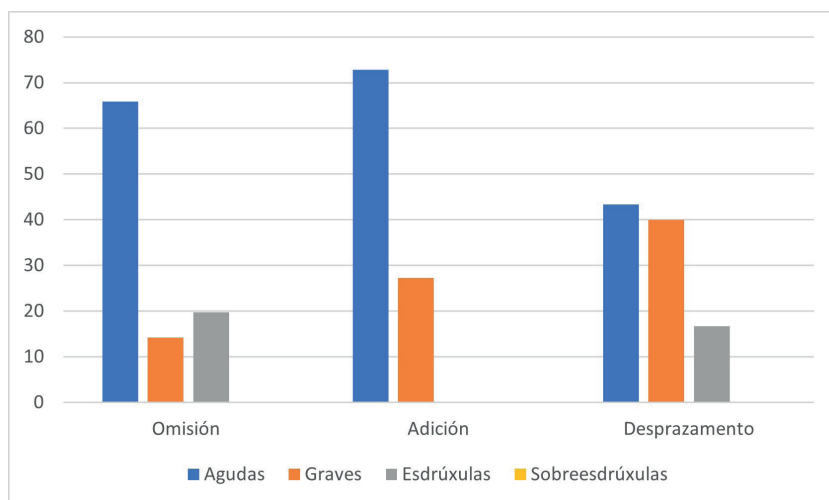
En segundo lugar, procedeuse á análise segundo a estrutura acentual da palabra. Na Gráfica 5 pode verse que o maior número de erros se concentra nas palabras agudas, mentres que graves e esdrúxulas se manteñen nuns valores moi inferiores. As sobresdrúxulas, pola súa parte, case non están presentes na mostra, de acordo coa súa escasa prevalencia na lingua.

GRÁFICA 5. Distribución dos erros segundo a estrutura acentual da palabra afectada



Para poder interpretar estes datos cómpre examinar como se combinan os dous principais parámetros empregados para a clasificación dos erros da mostra: o tipo de erro de colocación do acento gráfico e a afectación segundo a estrutura acentual da palabra. Así, na Gráfica 6 plásmanse as frecuencias relativas dos erros de colocación do acento ordenados tendo en conta a estrutura acentual da palabra.

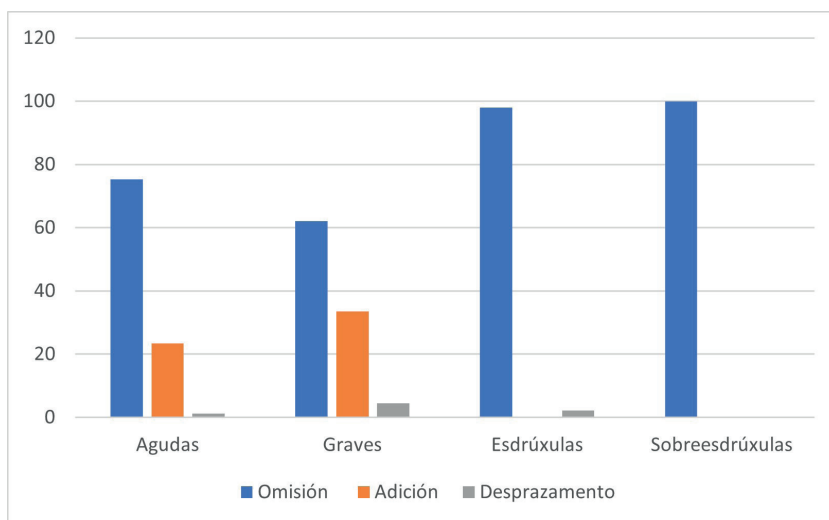
GRÁFICA 6. Incidencia dos erros de colocación segundo a estrutura acentual



Pode comprobarse que as agudas manifestan unha maior incidencia nas tres dinámicas. As graves non ocupan o primeiro lugar en ningún dos parámetros, aínda que contan cunha presenza importante na mostra, como corresponde ao padrón acentual dominante do galego. Loxicamente, non se deron casos de palabras esdrúxulas nas adicións, pois ao requirir sempre o acento estas palabras só son susceptibles de erro por omisión ou desprazamento. Nas sobresdrúxulas a única tipoloxía de erro rexistrada é a omisión, aínda que non se reflicten na gráfica pola súa escasa frecuencia (0,13%). Ademais, as altas porcentaxes das agudas confirman os resultados vertidos na Gráfica 5. Por último, cómpre salientar que os desprazamentos presentan unha diferenza mínima entre agudas e graves.

Na Gráfica 7 plásmase a incidencia dos erros segundo a estrutura acentual da palabra, ordenados polo tipo de fenómeno de colocación do acento. Detéctase a preponderancia das omisións en todas as clases de palabra, o que de novo corrobora os datos da Gráfica 1. Só nas graves a adición acada unha certa presenza, do 33%. Tamén se pode ver que o desprazamento amosa unha incidencia moi escasa e que no caso das sobresdrúxulas, como queda dito, só se rexistran omisións.

GRÁFICA 7. Incidencia dos erros segundo a estrutura acentual da palabra ordenados polo tipo de fenómeno de colocación do acento



Estas gráficas evidencian unha serie de tendencias: o erro cuantitativamente máis frecuente é a omisión; as agudas son as palabras nas que se concentra o maior número de erros; a omisión resulta o fenómeno dominante en todas

as clases de palabra; e as agudas vense máis afectadas polos tres fenómenos de acentuación.

Para interpretar axeitadamente estes datos cómpre correlacionalos coas tendencias xerais da lingua referentes á frecuencia das estruturas silábicas e acentuais e coa propia lóxica das regras de acentuación. A conformación mesma das regras de acentuación persegue marcar a tonicidade silábica co menor número posible de acentos gráficos, como sinala Scliar-Cabral (2020, p. 14) para o portugués² ou Medina Guerra (1995, p. 35) para o castelán, aínda que sempre persistan puntos do sistema que se resisten á sistematización (Perlin, 2007). Pode dicirse que as regras de acentuación nas linguas nas que se empregan procedementos para a marcación da tonicidade se rexen por principios ilustrados de racionalidade, sinxeleza e utilidade (Oroño, 2016); non en van a propia creación de moitas das academias que iniciaron os procedementos de prescrición ortográfica produciuse no marco do contexto ilustrado. Así, se ben a economía lingüística é un principio intrínseco á natureza da propia linguaxe (Paredes Duarte, 2010), tamén se aplica naqueles puntos do sistema que non dependen da acción espontánea da comunidade lingüística senón das decisións conscientes dos lingüistas. Para conseguir a marcación da tonicidade desde un principio de economía lingüística, correlaciónanse dous parámetros, tendo en conta a súa frecuencia na lingua. Tomaremos como referencia os estudos de Antonio Quilis (1981, 1983) a partir dun corpus de máis de 20.000 palabras para establecer os padróns acentuais maioritarios, se ben os datos non son directamente extrapolables ao galego, no que certos padróns incrementarían o número de palabras agudas en relación co castelán (*camións-camiones; pasteis-pasteles...*). Con todo, a semellanza no comportamento silábico e acentual entre as dúas linguas fai posible tomar os datos obtidos por Quilis como punto de referencia das pautas xerais:

² «Proclamada a República Portuguesa, em 5 de outubro de 1910, a regulamentação da ortografia foi definida por uma Comissão Ortográfica, sendo Gonçalves Viana seu Relator. Profundo conhecedor do português, aplicou o seguinte princípio de economia ao sistema de escrita: as palavras com acento de intensidade, mais frequentes, não levariam acento gráfico» (Scliar-Cabral, 2020, p. 14).

FIGURA 1. Frecuencia de aparición dos padróns acentuais do español. Fonte: Quilis (1983)

Frecuencia de aparición de los patrones acentuales del español

Esquema	Frecuencia absoluta	Frecuencia relativa
Oxítono	1.629	17,68 %
Paroxítono	7.336	79,50 %
Proparoxítono	254	2,76 %
Totales	9.219	99,94 %

Palabras	Frecuencia relativa
Paroxítonas	30,01 %
Monoslabas átonas	32,99 %
Monoslabas tónicas	17,59 %
Oxítonas	7,98 %
Bislabas átonas	3,57 %
Proparoxítonas	1,22 %
Adverbios en <i>-mente</i>	0,57 %

Por unha banda, as palabras esdrúxulas e sobreesdrúxulas, se excluimos as que xorden contextualmente en galego como resultado de engadir clíticos ás formas verbais, son sempre cultismos desde o punto de vista diacrónico que non experimentaron todas as evolucións patrimoniais. Son claramente as máis escasas no idioma, e por iso se marcan sempre mediante acento gráfico. Aplícase o mesmo procedemento ca no portugués ou no castelán e pola mesma razón de frecuencia: «Aplicando o principio da economía, por serem as mais raras, todas as proparoxítonas levam acento gráfico, como em ‘proparoxítonas’» (Scliar-Cabral, 2020, p. 19). Resta, pois, procurar un mecanismo para distinguir dous padróns acentuais: as palabras oxítonas e paroxítonas. No galego hai un predominio das palabras graves, como acontece no portugués (Scliar-Cabral, 2020, p. 15) e no castelán (Quilis, 1983; Figueroa de Amorós, 1994, p. 92). Combinando o criterio de frecuencia destes dous padróns acentuais co dos fonemas máis recorrentes en posición final de palabra (as vogais, dada a abundancia de sílabas libres no idioma, e as consoantes /n/ e /s/, altamente recorrentes e, no segundo caso, implicada como morfema de número ou desinencia verbal) chégase a unha norma que pon o acento gráfico nas palabras coa distribución acentual máis prevalente no

idioma cando esta circunstancia conflúe co peche silábico menos frecuente e, pola contra, nas que implican un menor número de elementos léxicos cando teñen os peches silábicos máis frecuentes³. O resultado, como queda dito, é que a tonicidade é deducible co menor número posible de acentos gráficos. De feito, algunha das diferenzas notables entre as regras de acentuación do galego e do castelán obedece a esta mesma pauta de minimizar o número de acentos precisos. Así, a excepción que se aplica en galego á acentuación das palabras agudas, de xeito que non levan acento gráfico cando presentan na sílaba tónica un ditongo decrecente, obedece á alta frecuencia deste padrón na lingua galega, mentres que en castelán practicamente se limita a certas formas verbais (2ª persoa de plural do presente de indicativo dos verbos da 1ª e 2ª conxugación), polo que as normas non formulan semellante excepción.

Así, mentres as palabras graves, as máis frecuentes en galego, levan acento gráfico cando a palabra remata coas terminacións menos frecuentes, a situación invértese no caso das agudas, que xustamente levan acento gráfico cando rematan nas terminacións máis frecuentes. O resultado é que a porcentaxe de palabras agudas que se acentúan graficamente é moi superior á das palabras graves, o que deixa como resultado lóxico unha alta frecuencia da omisión e de adición cando a sílaba tónica é a última da palabra. Con todo, a análise estatística revela a importancia de incidir, no tratamento didáctico, na acentuación das palabras agudas, por concentrar estas o número máis elevado de erros totais.

4. ANÁLISE DA DISPERSIÓN DOS ERROS

Outro parámetro importante que se debe considerar á hora de traballar cos datos é a súa distribución. Deste xeito, pode determinarse se os erros de acentuación resultan unha tendencia xeral en todo o alumnado, ou se, polo contrario, estes se concentran nuns poucos estudantes.

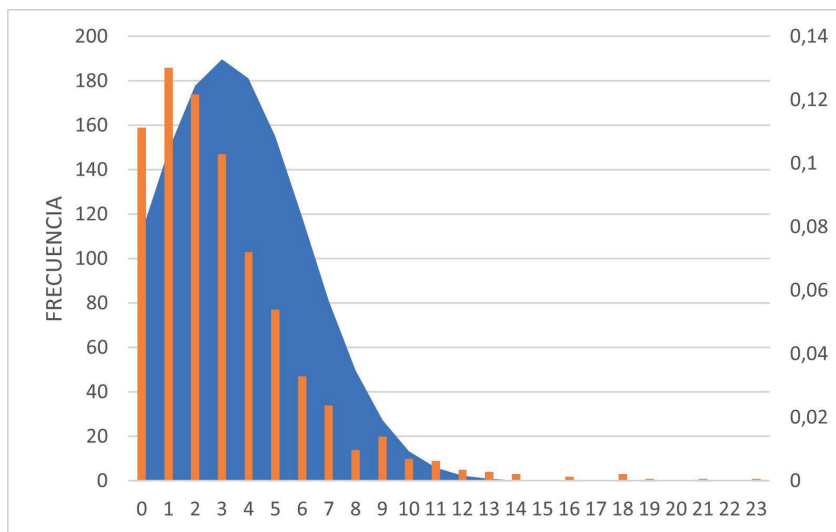
Xa se comentou que os problemas de acentuación se definen coma unha das dificultades do alumnado que máis se rexistra en *CORTEGAL*. O corpus

³ Hernando Cuadrado (2015) realiza unha revisión exhaustiva das sucesivas normas e modificacións que foi experimentando a acentuación gráfica en castelán e os motivos e argumentos que as foron sustentando, sinalando na súa conformación final a operatividade deste principio de economía: «Desde entón, en los sucesivos tratados ortográficos de la Real Academia Española se han ido formulando y revisando las reglas de acentuación gráfica de las palabras «teniendo en cuenta el principio de economía, es decir, sin necesidad de señalarla de manera explícita en todos los casos, sino solo en aquellos en que se considere preciso», de manera que las palabras ajustadas al patrón prosódico acentual más común no llevan tilde, mientras que aquellas otras que se encuentran situadas al margen de dicha pauta se marcan gráficamente» (Hernando Cuadrado, 2015, p. 153).

está composto por un total de 1000 exames, dos cales 841 contan con etiquetas relativas á acentuación, o que se traduce nunha porcentaxe do 84,1%.

Polo tanto, os erros de acentuación están presentes na maior parte do alumnado da mostra, mais convén analizar a súa distribución para comprobar o seu grao de dispersión. Se temos en conta os valores estatísticos, a media de erros de acentuación por exame respecto ao total de exames de *CORTEGAL* sitúase en 3,086 e a desviación estándar chega a 3. Trátase pois dun fenómeno frecuente na mostra, cun número medio de erros significativo se temos en conta a extensión dos textos (entre 200 e 250 palabras), e considerando a desviación estándar, non existe unha gran dispersión, o que se plasma na Gráfica 8:

GRÁFICA 8. Dispersión e frecuencia dos erros de acentuación por exame respecto ao total de exames de *CORTEGAL*



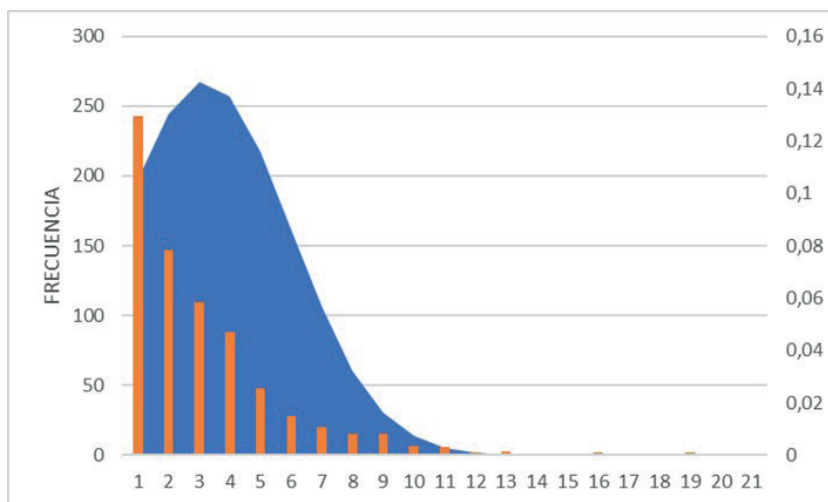
A maior parte dos valores concéntranse, pois, moi preto da media e a cantidade máxima de erros rexistrados é 23. A maioría dos exames presentan entre 1 e 4 erros de acentuación, posto que son estes valores os que superan a frecuencia de 100. Pola contra, os exames con máis de dez erros atópanse con frecuencias por debaixo de 10. Cabe ter en conta que un total de 159 probas de *CORTEGAL* non manifestan ningún erro.

Por outra banda, resulta interesante comprobar como se distribúen os erros de omisión e os de adición. Dos 1000 exames de *CORTEGAL*, 740 posúen etiquetas relativas á omisión de acento gráfico e 419 contan con

etiquetas de adición. Isto tradúcese en que os fenómenos de adición teñen presenza nun 41,9% dos exames e os de omisión nun 74%.

Se observamos a distribución dos erros de omisión, a media é de 3,19 por exame e a desviación estándar é de 2,79. A media resulta moi semellante á do número de erros de acentuación por exame e a dispersión tamén é escasa, tal e como recolle a Gráfica 9:

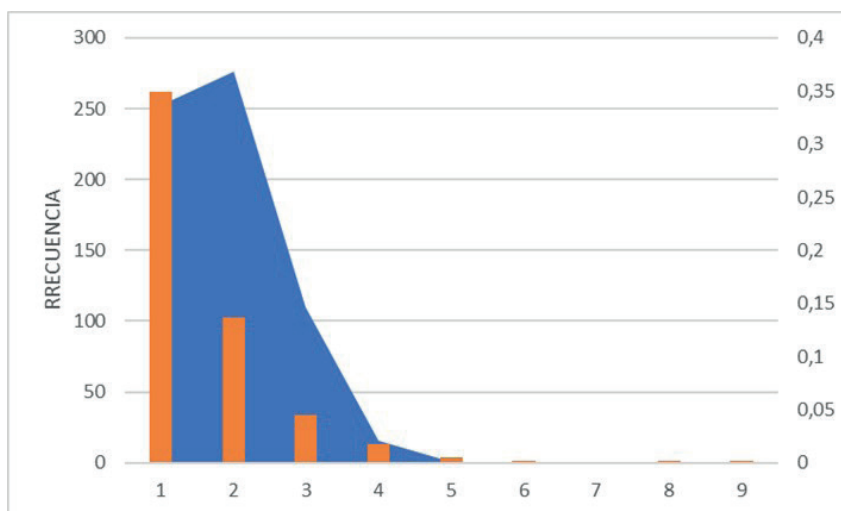
GRÁFICA 9. Dispersión e frecuencia dos erros de omisión por exame



O rango do número de erros rexistrados abrangue ata os 21 e os valores concéntranse moi preto da media. Consecuentemente, na maioría dos exames detéctanse entre 1 e 5 erros e as probas con máis de dez erros móvense en frecuencias por debaixo de 10. Os resultados amosan que as dificultades relativas á omisión de acento gráfico se presentan na maior parte do alumnado que comete erros de acentuación.

No caso das adicións, a media resulta de 1,59 e a desviación estándar de 0,99. Estes valores implican, de novo, un nivel moi escaso de dispersión, menor que o das omisións, como indica a Gráfica 10:

GRÁFICA 10. Distribución e frecuencia dos erros de adición por exame



Nos exames só se etiquetaron entre 0 e 9 fenómenos de adición. Ademais, a gran maioría das probas conta cunha ou dúas etiquetas de adición. Isto significa que preto da metade do alumnado da mostra de *CORTEGAL* que presenta erros de acentuación comete un ou dous erros de adición, superando un número limitado os 3 erros.

A análise da dispersión revela, pois, que a maioría do alumnado galego, nunha porcentaxe do 84,1%, comete erros de acentuación, e que se ben son poucos os que se sitúan no extremo dun número moi elevado de erros, nun texto de 200-250 palabras, nun contexto de alta motivación como son as probas ABAU, a maioría do alumnado comete entre 1 e 5 erros de acentuación, cunha media de case 3,09, o que apoia a necesidade de reformular as metodoloxías didácticas empregadas.

5. CONCLUSIÓNS E IMPLICACIÓNS DIDÁCTICAS

A análise dos erros de acentuación nos mil textos que conforman o corpus *CORTEGAL* permite ter unha visión das dificultades do alumnado galego coa acentuación na lingua propia. Os erros rexístranse nunha porcentaxe elevada do alumnado que realizou as probas ABAU tras a finalización do bacharelato (84,1%) e nunha proporción importante (81,79%) estes desvíos da norma son atribuíbles á interferencia entre os sistemas de acentuación do galego e do castelán. A grande prevalencia de erros de adición de acento en adverbios rematados en *-mente* ou en formas como **ademáis* ou **quizáis*, así como

en interrogativos directos e indirectos, demostrativos e pronomes e outras formas con acentuación diacrítica diferencial entre as dúas linguas, así como a dificultade manifesta coas formas verbais con clítico propias do galego revelan a confusión entre os sistemas acentuais, unha indistinción das normas que regulan ambos os idiomas que, sen dúbida, non é privativa da acentuación senón extensible a outros aspectos da norma lingüística. Neste sentido, a análise dos datos apoia a necesidade dun tratamento integrado de linguas que poña o foco nos aspectos diferenciais dos sistemas lingüísticos, unha reflexión explícita sobre os puntos nos que ambos os sistemas difiren para que haxa unha conciencia clara das diferenzas que minimicen as interferencias.

Unha análise semellante á realizada en *CORTEGAL* cos exames en castelán podería iluminar en que medida as interferencias funcionan na dobre dirección e se inversamente ao observado nos exames en galego os erros de acentuación en castelán deitan resultados semellantes desde o punto de vista cuantitativo e son explicables polos aspectos diferenciais co galego. Para corrixir a situación cómpre pór o foco neses momentos iniciais da aprendizaxe da norma e na necesidade de que a aprendizaxe desde o galego estea presente desde os primeiros achegamentos ao sistema acentual. No canto de iniciarse na acentuación no castelán e logo trasladala ao galego marcando as diferenzas nesta lingua, por efectividade na aprendizaxe de ambos os códigos así como polo que se transmite implicitamente sobre as linguas, a aprendizaxe debería ser simultánea, a través dun tratamento integrado, e non, como adoita acontecer neste e nos demais aspectos da aprendizaxe lingüística, maioritariamente desde o castelán e secundariamente e diferencialmente para o galego, o cal transmite a idea dunha norma-base para o castelán e un tratamento por desvío ou diferenza para o galego. Non en van no conxunto do estado foi desde os ámbitos con lingua propia e dúas linguas ambientais onde se prestou maior atención ao Tratamento Integrado de Linguas (TIL) (Montagut e Pereña, 2018; Martí Climent, 2019; Ibarluzea Santisteban et al., 2021). Así, Elorza e Muñoa (2008) ven no TIL unha ferramenta para a promoción das linguas minorizadas e o mesmo afirma Rama Lorenzo (2020) para o galego.

Entre os piares do TIL, Guasch (2008) incorpora as actitudes cara ás linguas no que hoxe poderíamos chamar unha actitude ecolingüística e tamén a imbricación da reflexión interlingüística e o uso da lingua. Este último aspecto, o da reflexión interlingüística, revélase especialmente central para a cuestión da acentuación, como o é para o conxunto do enfoque do TIL (Ruiz Bikandi, 2008). Cando se realiza o ensino e aprendizaxe das normas de

acentuación, en poucas ocasións se tira proveito do seu potencial para a reflexión metalingüística. Non é común que os materiais curriculares acompañen as normas de acentuación e as actividades de aplicación dunha reflexión sobre o sentido e a razón da formulación destas. Malia que nos currículos se adopten termos como «coñecemento da lingua» ou «reflexión sobre a lingua» no canto de «gramática», poucas veces se incorpora, nin nos materiais curriculares nin nas prácticas de aula, a reflexión sobre o sentido e a orixe das propias normas. As regras de acentuación, pola contra, constitúen un punto da norma lingüística idóneo para reflexionar sobre a normativa lingüística, a súa funcionalidade e a súa inspiración en principios como a economía lingüística, a función de discriminación ou a procura do axuste á realidade fonolóxica. Incorporar esta perspectiva reflexiva sobre as normas lingüísticas permitiría comprender, ademais, a razón última das diferenzas entre linguas. Así, no tocante á acentuación dos adverbios en *-mente*, a norma do castelán introduce criterios diacrónicos, axustándose a do galego á realidade sincrónica actual da lingua; a excepción aplicada á acentuación das palabras agudas conecta cos criterios de economía e funcionalidade que rexen as regras de acentuación, dada a alta prevalencia deste tipo de palabras no galego; a norma de acentuación do castelán para as secuencias *-iu-*, *-ui-* é doada de aplicar porque renuncia a axustarse á realidade fonética, mentres que a norma do galego si discrimina entre que estas secuencias conformen ditongos homoxéneos ou hiatos; o afondamento na acentuación diacrítica, pola súa banda, permite reflexionar sobre as diferentes coincidencias que en ambas as linguas se dan entre formas con distinto valor gramatical.

Este campo de aprendizaxe da norma lingüística para o galego e o castelán préstase pois a potenciar a reflexión interlingüística, á que pode chegarse a través da técnica de formulación de preguntas eficaces (López Valero et al., 2020). Partindo de constatar, nunha fonte de autoridade ortográfica como pode ser o dicionario, discrepancias na acentuación, pódense formular preguntas como: por que *construír* se acentúa en galego e non en castelán?, por que *rapidamente* non se acentúa en galego e si en castelán?, por que *ademais* non se acentúa en galego e *amáis* si se acentúa en castelán? Mais a reflexión non debe limitarse ao funcionamento distintivo da norma, senón profundar no sentido último das normas de acentuación, na procura que se fixo dun sistema de marcación da tonicidade rexido por principios de economía, eficacia e discriminación lingüística.

En conclusión, o tratamento didáctico da acentuación precisa de axustes e cambios se se pretende mellorar a execución e corrección e minimizar as

interferencias. Un achegamento inicial desde a perspectiva do TIL que non se limite a transmitir unhas regras e promover a súa aplicación mecánica, senón que se sirva deste punto do sistema especialmente sistemático, con relevancia fonolóxica e con función de discriminación entre formas lingüísticas, para promover unha actitude reflexiva cara aos fenómenos lingüísticos non só melloraría a execución e minimizaría as interferencias, senón que tería un valor na transmisión de actitudes lingüísticas positivas cara aos diversos códigos acorde cos principios de defensa do plurilingüismo recollidos na *Carta Europea das Linguas Rexionais e Minoritarias* (Consello de Europa, 1992) que urxe promover se se quere verdadeiramente apostar por unha Europa plurilingüe e comprometida coa súa diversidade lingüística e cultural.

REFERENCIAS BIBLIOGRÁFICAS

- ÁLVAREZ, B. (1977). Los errores ortográficos más frecuentes en composiciones escritas por niños de primero y segundo años de primaria en México. *Revista del Centro de Estudios Educativos*, 7(4), 92-97.
- CONSELLO DE EUROPA (1992). *Carta Europea das Linguas Rexionais e Minoritarias*. Consello de Europa.
- ELORZA, I. e MUÑOA, I. (2008). Promoting the minority language through integrated plurilingual language planning: The case of the ikastolas. *Language, Culture and Curriculum*, 21(1), 85-101. <https://doi.org/10.2167/lcc345.0>
- FIGUEROA DE AMORÓS, E. (1994). La acentuación ortográfica en la educación superior. *Educación*, 3(5), 81-93. <https://doi.org/10.18800/educacion.199401.005>
- GONZÁLEZ PASCUAL, B. (2003). Problemas de acentuación en el español actual. *Interlingüística*, 14, 489-496.
- GUASCH, O. (2008). Reflexión interlingüística y enseñanza integrada de lenguas. *Textos de didáctica de la lengua y la literatura*, 47, 20-32.
- HERNANDO CUADRADO, L. A. (2015). Acento prosódico y acentuación gráfica en español. *Archivum. Revista de la Facultad de Filosofía y Letras*, 65, 133-164. <https://doi.org/10.17811/ARC.65.2015.133-164>
- IBARLUZEA SANTISTEBAN, M., ZULOAGA SAN ROMÁN, E., BOILLOS PEREIRA, M. M. e ARANBERRI MONASTERIO, N. (2021). La implementación del tratamiento integrado de lenguas: percepciones del profesorado de educación primaria y secundaria. *Tonos digital. Revista de estudios filológicos*, 40. <http://hdl.handle.net/10201/105162>
- LÓPEZ VALERO, A., VALVERDE CARAVACA, R. e ENCABO FERNÁNDEZ, E. (2020). Una investigación sobre el Aprendizaje Integrado de contenidos y Lenguas extranjeras (AICLE) en Educación Primaria. El uso de técnicas para la formulación de preguntas eficaces. *Porta Linguarum. Revista internacional de didáctica de las lenguas extranjeras*, 33, 63-74. <https://doi.org/10.30827/portalin.vi33.18125>

- MARTÍ CLIMENT, A. (2019). El tratamiento integrado de lenguas en la gramática: una aproximación a la terminología básica en los libros de texto de secundaria. *ReGroc. Revista de gramática orientada a las competencias*, 2(1), 89-116. <https://doi.org/10.5565/rev/regroc.44>
- MEDINA GUERRA, A. M. (1995). Voces con doble acentuación en el diccionario de la Academia. *REALE. Revista de Estudios de Adquisición de la Lengua Española*, 4, 35-66. <http://hdl.handle.net/10017/7353>
- MONTAGUT, M. e PEREÑA, M. (2018). Enseñar lenguas desde un enfoque metodológico global. *Textos de didáctica de la lengua y la literatura*, 81, 43-49.
- OROÑO, M. (2016). La polémica sobre acentuación ortográfica entre Carlos Martínez Vigil y Fidelis del Solar. *Nueva Revista del Pacífico*, 64, 67-96. <https://doi.org/10.4067/S0719-51762016000100003>
- PAREDES DUARTE, M. J. (2008). El principio de economía lingüística. *Pragmalingüística*, 15-16, 166-178.
- PERLIN, J. (2007). Las inconsecuencias de la acentuación española. *Itinerarios. Revista de estudios lingüísticos, literarios, históricos y antropológicos*, 5, 183-188.
- POLO, J. (1974). *Ortografía y ciencia del lenguaje*. Paraninfo.
- QUILIS, A. (1981). *El acento español*. Universidad Nacional Autónoma de México.
- QUILIS, A. (1983). Frecuencia de los esquemas acentuales en español. En *Estudios ofrecidos a Emilio Alarcos Llorach* (pp. 113-126, vol. 5). Universidad de Oviedo.
- RAMA LORENZO, S. (2020). O Tratamento Integrado de Linguas como Instrumento para a normalización da lingua galega. *Revista galega de educación*, 78, 32-35.
- RUIZ BIKANDI, U. (2008). La reflexión metalingüística desde varias lenguas. *Textos de didáctica de la lengua y la literatura*, 47, 33-45.
- SCLIAR-CABRAL, L. (2020). Conhecimentos necessários para saber ler os acentos gráficos no português. *Confluência. Revista do Instituto de Língua Portuguesa*, 59, 9-24. <https://doi.org/10.18364/rc.v1i59.448>
- SOLLA PORTELA, M. A. (2011). Módulo de acentuación para o galego en Freeling. *Linguamática*, 2(3), 59-64.
- SOTOMAYOR, C., ÁVILA, N., BEDWELL, P., DOMÍNGUEZ, A., GÓMEZ, G. e JÉLDREZ, E. (2017). Desempeño ortográfico de estudiantes chilenos: claves para la enseñanza de la ortografía. *Estudios Pedagógicos*, 43(2), 315-332. <https://doi.org/10.4067/S0718-07052017000200017>
- TOMCSÁNYI MAJOR, J. (2015). Hipercorrección ortográfica en la marcación del acento en documentos académicos y divulgativos en Costa Rica. *Letras*, 58, 29-49. <https://doi.org/10.15359/rl.2-58.2>

O léxico non estándar en *CORTEGAL*

MARÍA ÁLVAREZ DE LA GRANJA

Instituto da Lingua Galega-Universidade de Santiago de Compostela

1. INTRODUCCIÓN

Algúns traballos previos realizados con estudantado galego da educación secundaria (López Meirama e Álvarez de la Granja, 2014, p. 60; Loredó e Silva Valdivia, 2020, p. 143) mostran a existencia de deficiencias no dominio do código galego estándar no ámbito léxico. Por tal motivo, non resulta estraño que o corpus *CORTEGAL*, presentado en Álvarez de la Granja e Rodríguez Rodríguez neste volume, recolla un considerable número de formas léxicas non estándar. O obxectivo deste capítulo é ofrecer unha análise, cuantitativa e cualitativa, de tales formas. Traballamos cos datos obtidos a través do buscador do corpus, que nos permite localizar todas as formas etiquetadas con códigos do nivel léxico e que nos ofrece unha primeira clasificación dos resultados, clasificación que é comentada e completada ao longo do traballo.

Ofrecemos en primeiro lugar, no apartado §2, unha presentación do sistema de clasificación das formas léxicas non estándar empregado en *CORTEGAL*. A seguir, no apartado §3, amosamos os resultados obtidos. Despois dunha presentación dos datos globais, analizamos máis detalladamente as unidades léxicas non estándar, organizadas de acordo coas diferentes clases establecidas (formas léxicas con diferenzas de acentuación con respecto á voz normativa, con diferenzas de xénero e con diferenzas que van máis alá do xénero e da acentuación). O subapartado dedicado ao último grupo, o máis numeroso dos tres mencionados, estrutúrase atendendo á orixe das formas non estándar. Ademais, neste apartado §3 ofrecemos información que delimita máis axeitadamente cada unha das clases, explicadas en §2 de maneira

xeral. A seguir, no apartado §4 presentamos as conclusións principais, discutimos os datos atopados desde a perspectiva do uso do repertorio lingüístico das/dos falantes multilingües así como do fenómeno do contacto, e reflexionamos tamén sobre a utilidade e limitacións de *CORTEGAL* para extraer conclusións ao respecto do coñecemento que ten o estudantado galego sobre o código galego estándar no nivel léxico. O capítulo péchase cun Apéndice onde se recollen as voces presentadas ao longo do traballo con información sobre a súa presenza noutras fontes.

2. CLASIFICACIÓN DAS FORMAS LÉXICAS NON ESTÁNDAR

No nivel léxico de *CORTEGAL* estandarízanse e codifícanse todas aquelas palabras ou todas aquelas expresións pluriverbais fixas que non forman parte do código normativo da lingua galega. Concretamente, son tres os tipos de formas léxicas non estándar establecidos, que se manteñen diferenciados a través de cadansúas etiquetas e que a seguir se presentan ordenados de acordo coa súa relevancia cuantitativa. Nos apartados correspondentes, ofreceremos máis pormenores sobre cada unha das clases propostas:

- substitución dunha palabra ou expresión estándar por unha unidade léxica non estándar con diferenzas que van máis alá do xénero ou da acentuación, por exemplo, o emprego de *plato* en vez de *prato* (código L_w_su).
- atribución a un substantivo dun xénero non estándar, por exemplo, a utilización do substantivo masculino *leite* como feminino: «a leite» (código L_gen_su).
- colocación do acento de intensidade nunha sílaba diferente da acentuada na forma estándar, por exemplo, o emprego de *Nóbel* no canto de *Nobel* (código L_ac_su).

Por outro lado, na maior parte dos casos tamén se indica a orixe da forma non estándar organizada a través das seguintes subclases, que se presentan co mesmo criterio ordenador:

- palabras non estándar que coinciden co español e teñen a súa orixe nesta lingua, por exemplo, *salir* (código L_sp).
- palabras non estándar que resultan de calcos a partir de formas do español, por exemplo, *á longa* a partir de *a la larga*; de adaptacións de voces do español, por exemplo, *antoxo*, a partir de *antojo*; ou de mesturas de difícil sistematización entre segmentos fónicos da palabra galega e da

- palabra castelá, por exemplo, *ventaxe* [español *ventaja*, galego *vantaxe*] (código L_spadapt).
- palabras, como *almacear*, tradicionalmente denominadas hipergaleguismos ou hiperenxebrismos, en que se sobrexeneraliza un padrón de equivalencias galego-castelán, pero nos que a forma ou secuencia adaptada (neste caso *almacenar*) é estándar e coincidente en galego e castelán (código L_hc).
 - formas que se crean por analoxía con outras voces galegas, como *consomer*, no canto de *consumir* (probablemente a partir das formas estándar do presente de indicativo *consume*, *consoles* e *consomen*) ou *lalinés* en vez de *lalinense* (sobre *vigués*, *cangués*...) (código L_anal).
 - formas tiradas do galego popular ou do galego elaborado (empregando os termos de Dubert, 2005)¹ e que non responden a ningunha das orixes anteriores, como a variante dialectal *bon* ou a forma *voltar* ‘volver’, frecuente no galego elaborado probablemente por transferencia desde o portugués (código L_gal).
 - formas que proceden da mestura e confusión entre dúas formas ou expresións: por exemplo, *a saber Deus*, probable mestura de *a saber* e *sabe Deus*, ou o substantivo *indixesta*, que resulta moi posiblemente do cruzamento entre o substantivo *inxesta* e o adxectivo *indixesto* (código L_mix).
 - formas procedentes doutras linguas (excluído o castelán) e que, de acordo coa proposta normativa do galego, deberían adaptarse, por exemplo, *tablet* en vez de *tableta* (código L_for).

3. ANÁLISE DAS FORMAS LÉXICAS NON ESTÁNDAR

Neste apartado presentaremos os principais resultados atopados: en §3.1 farémolo de maneira xeral e en §3.2 estableceremos subgrupos de acordo cos tipos de diverxencia con respecto ao estándar sinalados e coas súas orixes. Pero antes de tal presentación cómpre facer dúas precisións:

En primeiro lugar, non é descartable que algunhas das formas etiquetadas e computadas como formas non estándar sexan en realidade o resultado dun lapsus cálimi (sobre esta cuestión, *vid.* §3.2.1). Por exemplo, *dineiro* pode responder simplemente a un esquecemento do til sobre a nasal ou *ver-sonar* a un erro de escritura que levou á omisión do <i>. Por tal motivo, cómpre tomar con cautela algunhas das formas proporcionadas, especialmente aquelas das que só temos un exemplo.

¹ «Una variedad es el gallego culto o elaborado, que aparece en los enunciados orales y escritos más formales o institucionales; la otra es el gallego popular, vernáculo o no elaborado, que aparece en enunciados generalmente hablados, más alejados del mundo oficial» (Dubert, 2005, p. 274).

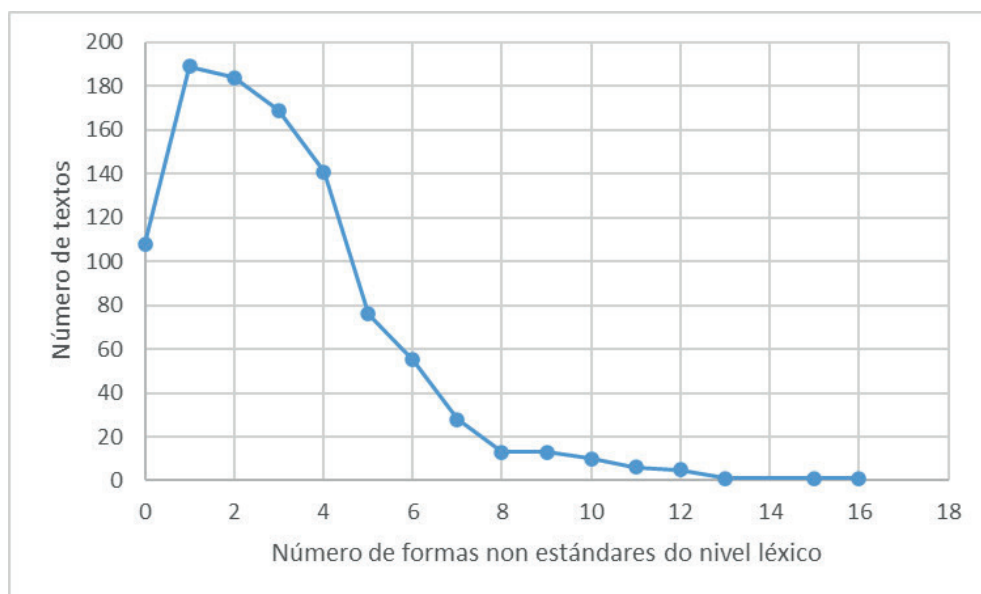
En segundo lugar, as palabras son corrixidas ortograficamente e presentadas, na maior parte dos casos, na forma de cita habitual de cada clase de palabras, pois só indicaremos a forma concreta flexionada cando sexa relevante para a análise.

3.1. Análise global

O número total de formas ou expresións non estándar é de 3025 e, de acordo co que se indica no capítulo de Álvarez de la Granja e Rodríguez Rodríguez (neste volume, §3.2), o número total de *tokens* do corpus é de 268.595. Se eliminamos os 25.004 signos de puntuación e os 193 números arábigos ou romanos, o resultado é que un 1,24% de *tokens* do corpus (palabras ou expresións²) son unidades léxicas non estándar. O promedio de palabras non estándar é de 3,03 por texto (lémbrese que estes teñen unha extensión recomendada de entre 200 e 250 palabras, *vid.* Álvarez de la Granja e Rodríguez Rodríguez neste volume, §2.1), cunha desviación estándar de 2,41 e valores que se sitúan entre 16 nun único texto e 0 en 108. Este rango é o que se recolle na Gráfica 1, que mostra a correlación entre o número de formas non estándar (eixe horizontal) e o número de textos que conteñen cada cifra de voces diverxentes (eixe vertical). Como se pode comprobar, o valor máis repetido ou moda é 1, con 189 textos que só teñen unha forma non estándar. A gráfica mostra ben ás claras como a medida que aumenta o número de voces diverxentes diminúe o número de textos. Esta tendencia só se rompe ao inicio, posto que a cantidade de textos con ningunha forma non estándar (108) é considerablemente inferior á cantidade de textos con só unha voz diverxente (189).

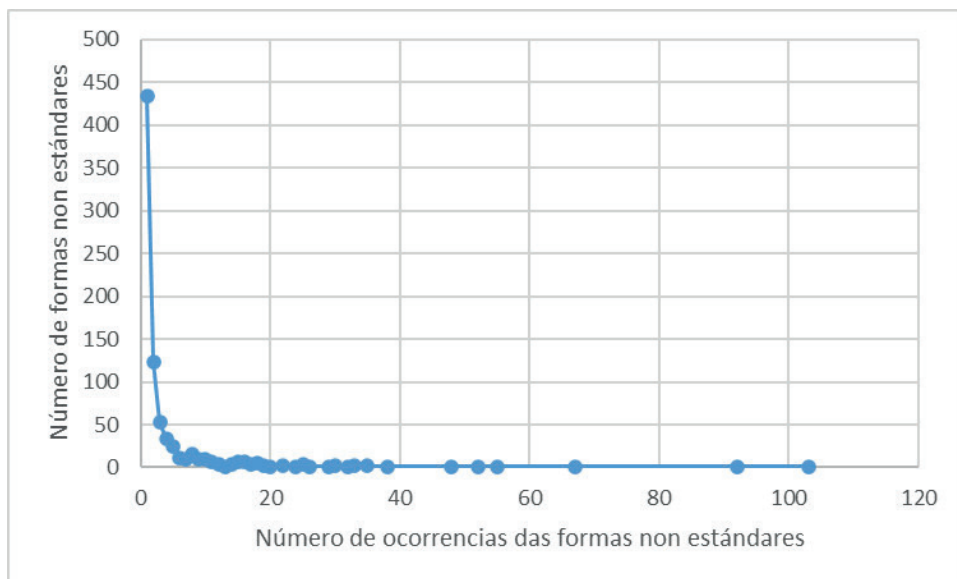
² Debe terse en conta que, tal e como se indica en Álvarez de la Granja e Rodríguez Rodríguez (neste volume, §2.5.4.), as expresións pluriverbais fixas só se unifican cando deben ser estandarizadas e etiquetadas conxuntamente. Por exemplo, unimos os elementos *sin* e *embargo* do castelanismo *sin embargo*, que pasa a computar como un único *token*, porque debe ser anotado e estandarizado como unha unidade. Outras formas pluriverbais, como *de vez en cando*, que non precisan de ningún tipo de anotación, non se unifican, de xeito que computan como varios *tokens*. Por outro lado, nas contraccións e nos verbos seguidos de pronome enclítico prodúcese a división nos compoñentes que os conforman, computando cada un deles como un *token* (*vid.* Álvarez de la Granja e Rodríguez Rodríguez neste volume, §2.2).

GRÁFICA 1. Correlación entre o número de formas non estándar do nivel léxico e o número de textos



As 3025 palabras ou expresións non estándar distribúense entre 787 formas distintas, o que supón unha media de 3,84 ocorrencias por forma, cunha desviación estándar de 7,93 e un rango que se sitúa entre 103 ocorrencias (no caso de *plato*) e 1 ocorrencia en 434 casos. A seguinte gráfica de dispersión (Gráfica 2) mostra a correlación entre o número de ocorrencias das formas non estándar nos textos de *CORTEGAL* (eixe horizontal) e a cantidade de formas non estándar con tales ocorrencias (eixe vertical). Como se ve claramente na imaxe, a moda ou o valor que máis se repite é 1, coas 434 formas xa indicadas, seguida de 124 formas con 2 ocorrencias. A tendencia continúa á baixa a medida que se incrementa o número de casos ata practicamente estabilizarse en 1 ou 2 formas a partir aproximadamente das 20 ocorrencias.

GRÁFICA 2. Correlación entre o número de ocorrencias das formas non estándar do nivel léxico e o número de formas non estándar



Por outro lado, o número total de códigos que identifican un problema léxico é de 3049, posto que unha expresión pode ter atribuída máis dunha etiqueta deste nivel³. A clase claramente máis numerosa é aquela constituída por formas con diferenzas con respecto ao estándar que van máis alá do xénero e da acentuación, que supón pouco menos do 95% dos casos, tal e como se ve na Táboa 1:

TÁBOA 1. Distribución das formas non estándar do nivel léxico segundo o aspecto diverxente do estándar

Clase	Código	Nº de códigos	Porcentaxe
Diferenzas que van máis alá do xénero e da acentuación	L_w_su	2887	94,69%
Diferenzas de xénero	L_gen_su	138	4,53%
Diferenzas de acentuación	L_ac_su	24	0,79%
Total		3049	

³ Sucede isto cando se acumulan dúas etiquetas diferentes (por exemplo, L_gen_su e L_w_su en *engranaxe*, palabra adaptada do español *engranaje*, empregada en masculino malia rematar no sufixo *-axe* e cuxa forma estándar correspondente é a voz feminina *engrenaxe*) ou cando se acumulan dúas etiquetas L_w_su (naqueles casos en que se emprega ou se calca unha expresión do castelán, que á súa vez contén un castelanismos no seu interior, como é o caso do conector *sin embargo*, que inclúe o castelanismos *sin*).

No que se refire á orixe das formas léxicas non estándar, os resultados son os que se ofrecen na Táboa 2, onde se observa claramente como as formas procedentes do español e coincidentes con esta lingua son as maioritarias. Se sumamos as formas con influencia directa ou indirecta do español (L_sp e L_spadapt), chégase practicamente ao 90% dos casos.

TÁBOA 2. Distribución das formas non estándar do nivel léxico de acordo coa súa orixe

Orixe	Código	Nº de códigos	Porcentaxe
Transferencias directas do castelán	L_sp	2270	71,77%
Transferencias do castelán con adaptación ou mestura co galego	L_spadapt	566	17,89%
Hipergaleguismos	L_hc	112	3,54%
Formas creadas por analogía a partir doutras voces galegas	L_anal	106	3,35%
Transferencias do galego popular ou do galego elaborado	L_gal	92	2,91%
Mestura ou confusión entre formas	L_mix	10	0,32%
Estranxeirismos non adaptados	L_for	7	0,22%
Total		3163	

Para interpretar estes resultados, debe terse en conta que, tal e como se indica en Álvarez de la Granja e Rodríguez Rodríguez (neste volume, §2.5.4), a clasificación en canto á orixe non se leva a cabo en todas as formas non estándar, senón só naquelas en que é posible establecer unha hipótese plausible que xustifique a presenza da forma e de acordo ademais cuns grupos restrinxidos e previamente fixados. Isto explica que en catro casos non se asigne unha etiqueta deste tipo (*vid.* §3.2.1.8). Ademais, cómpre sinalar que, tal e como se indica tamén no mesmo lugar do capítulo de Álvarez de la Granja e Rodríguez Rodríguez, en ocasións son ofrecidas varias hipóteses sobre a orixe dunha mesma forma non estándar: así, por exemplo, a palabra *cual* etiquétase como voz galega dialectal (L_gal) (*vid.* Fernández Rei, 1999, p. 51) e como transferencia do español (L_sp). Este feito explica que a suma de códigos de orixe sexa superior á suma de códigos ofrecidos na Táboa 1. Ademais, tal e como indicamos para as clases que identifican o tipo de diverxencia do estándar, unha mesma forma pode acumular desviacións de diferente natureza que actúan simultaneamente e que poden responder a diferentes orixes: por exemplo, a forma *emplataxe* («no emplataxe») está creada sobre

o castelanismo *emplatar* (transferencia directa do español, L_sp), presenta o xénero masculino propio das palabras castelás acabadas no sufixo *-aje* (transferencia directa do español, L_sp) e é o resultado dunha creación da/ do estudante a partir de regras produtivas do galego (analogía, L_anal). O afirmado neste parágrafo explica que unha mesma forma poida aparecer comentada en diferentes apartados ao longo do traballo

A continuación, na Táboa 3 e na Táboa 4, ofrecemos a distribución das formas léxicas non estándar por categoría gramatical⁴. Debe terse en conta que, para facilitar a análise, foron computadas conxuntamente todas aquelas expresións complexas que resultaron dun proceso de unificación de varios *tokens*, proceso derivado da necesidade de estandarizar a secuencia conxuntamente nalgún nivel de análise, tal e como explicamos na nota 2. Na Táboa 3 ofrecemos a porcentaxe de formas non estándar sobre o total de formas de cada categoría, ordenadas de maior a menor presenza relativa.

TÁBOA 3. Distribución por categorías gramaticais das formas léxicas non estándar (I)

	Formas do corpus	Formas léxicas non estándar	Porcentaxe de formas léxicas non estándar sobre as formas do corpus
Expresións complexas	302	123	40,73%
Interxeccións	10	3	30,00%
Substantivos	52406	1355	2,59%
Adxectivos	15945	344	2,16%
Verbos	40936	781	1,91%
Adverbios	15896	147	0,92%
Conxuncións	19420	103	0,53%
Numerais	402 ⁵	1	0,25%
Pronomes e determinantes	62744	118	0,19%
Preposicións	35337	49	0,14%
Total	243398	3025	1,24%

Como se pode observar, salvando tanto a clase das interxeccións, que alcanza un 30% de formas non estándar (feito que pode non resultar moi significativo pola escaseza de unidades desta categoría), como as expresións

⁴ Traballamos coas categorías gramaticais de nivel superior que establece o lematizador Freeling (vid. §3.2 en Álvarez de la Granja e Rodríguez Rodríguez neste volume), pero unificando pronomes e determinantes, dado que moitas veces coinciden na súa forma.

⁵ Só se computan os numerais escritos en letra.

complexas, cunha presenza do 40,73% (e que precisamente en varias ocasións son o resultado dun proceso de unificación derivado da necesidade de realizar unha estandarización conxunta no nivel léxico, o que explicaría a porcentaxe tan elevada), son claramente as tres clases gramaticais máis produtivas e abertas, por esta orde, substantivos, adxectivos e verbos, as categorías cunha porcentaxe maior de formas non estándar. O 81,98% das formas léxicas non estándar pertence a algunha destas clases, aínda que só representan un 44,90% do total de formas do corpus. Na Táboa 4 preséntanse os datos de distribución das formas non estándar entre as diferentes clases gramaticais, tendo en conta desta volta a porcentaxe que supoñen sobre o total de voces non normativas e establecendo unha comparación coa porcentaxe que supón cada categoría gramatical sobre o total de formas do corpus.

TÁBOA 4. Distribución por categorías gramaticais das formas léxicas non estándar (II)

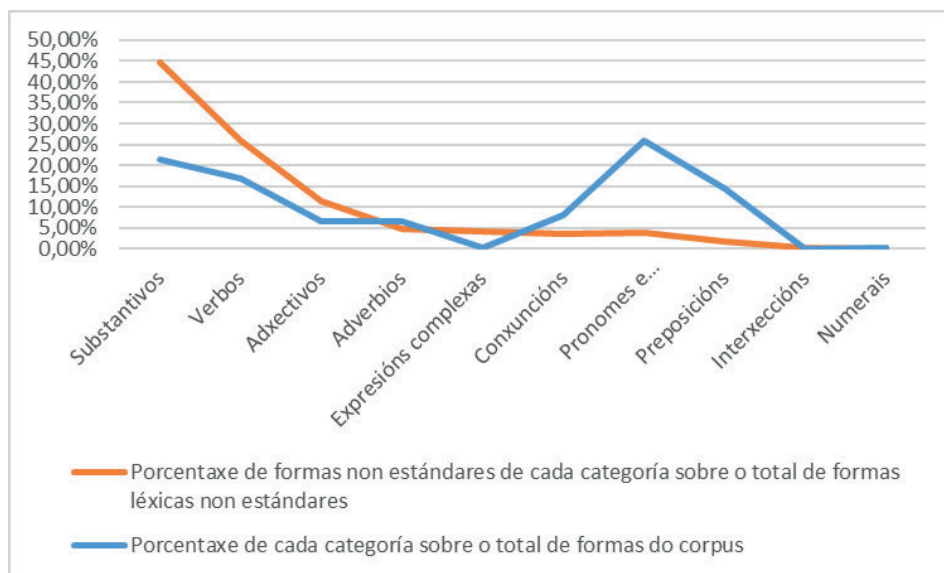
	Porcentaxe de cada categoría sobre o total de formas do corpus	Porcentaxe de formas non estándar de cada categoría sobre o total de formas léxicas non estándar
Substantivos	21,53%	44,79%
Verbos	16,82%	25,82%
Adxectivos	6,55%	11,37%
Adverbios	6,53%	4,86%
Expresións complexas	0,12%	4,07%
Conxuncións	7,98%	3,40%
Pronomes e determinantes	25,78%	3,90%
Preposicións	14,52%	1,62%
Interxeccións	0,004%	0,10%
Numerais	0,17%	0,03%

No caso das expresións complexas, substantivos, verbos, adxectivos e interxeccións, a porcentaxe de formas non estándar sobre o conxunto de palabras ou expresións non normativas é superior á porcentaxe de tales categorías no corpus, mentres que nos adverbios, preposicións, conxuncións, numerais, pronomes e determinantes sucede xustamente o contrario.

Na Gráfica 3 ilústrase esta mesma comparativa, que mostra claramente como é no grupo dos substantivos onde resulta especialmente significativa a presenza de formas non estándar. Pola contra, vese con claridade como é na

categoría de pronomes e determinantes onde a presenza é menor. Aínda que en sentido contrario, tanto nos substantivos como nos pronomes e determinantes a diferenza é de máis de 20 puntos con respecto ao nivel de presenza xeral no corpus de tales categorías.

GRÁFICA 3. Comparativa entre a distribución das categorías gramaticais no total do corpus e nas formas léxicas non estándar



3.2. Análise por clases

Despois de ofrecer os resultados globais, analizamos os datos organizados polo tipo de diverxencia con respecto ao estándar: diverxencias que van máis alá do xénero e a acentuación (§3.2.1), diverxencias no xénero (§3.2.2) e diverxencias na acentuación (§3.2.3). Con todo, antes de proceder á análise cómpre presentar o Apéndice que pecha o capítulo, posto que se farán varias referencias a el ao longo da exposición.

Entre as formas léxicas non estándar atopadas en *CORTEGAL* encontramos algunhas con cando menos certo grao de emprego no galego elaborado ou no galego popular (como por exemplo *sen embargo* ou *parexa*) ao lado doutras, como *inconvente* ou *innaturabilidade*, que non parecen rexistrarse en tales variedades. Agora ben, o número de ocorrencias dunha voz en *CORTEGAL* (e o número de textos en que aparece) non nos permite extraer conclusións sobre o seu grao de integración no galego cando nos atopamos con cifras baixas, posto que se trata dun corpus relativamente pequeno e cun

léxico moi condicionado pola temática dos textos (por exemplo, só hai unha ocorrencia de *corbata*, por moito que sexa unha voz frecuente no galego popular, pero debe terse en conta que o concepto ‘gravata’ só aparece unha vez en *CORTEGAL*). Ante este feito, consideramos de interese ofrecer en apéndice información sobre a presenza noutras fontes das diferentes formas que se proporcionan nas táboas que figuran ao longo do texto, o cal pode axudar a determinar o seu grao de difusión no galego e a avaliar máis adecuadamente o teor das voces rexistradas en *CORTEGAL*.

Concretamente, as fontes seleccionadas son o *Corpus de Referencia do Galego Actual (CORGA)* e o *Tesouro Informatizado da Lingua Galega (TILG)*, que sobre todo nos ofrecen información sobre o galego elaborado, así como o *Tesouro do léxico patrimonial galego e portugués (TLPGP)*, que nos proporciona datos sobre o galego popular. O *CORGA* é un corpus que recolle sobre todo textos escritos (especialmente pertencentes a libros e xornais) conformado, no momento en que levamos a cabo as consultas, por 43 millóns de palabras correspondentes ao período 1975-2020. O *TILG* é outro corpus, constituído sobre todo por textos escritos extraídos de libros, que contén ao redor de 30 millóns de formas dun período moito máis amplo, posto que inclúe textos que se sitúan entre 1612 e 2013⁶. Debe terse en conta, en calquera caso, que moitos dos textos que se inclúen nestes corpus, sobre todo os dos últimos anos, foron revisados por lingüistas, de maneira que é posible que algunhas creacións espontáneas e non normativas coincidentes coas rexistradas en *CORTEGAL* fosen suprimidas e substituídas por outras formas antes da súa publicación.

Finalmente, o *TLPGP* é unha ferramenta de distinto tipo, que permite acceder á información contida en diferentes obras, sobre todo de carácter lexicográfico. Estas obras, publicadas entre 1929 e a actualidade, reúnen léxico patrimonial recollido a través de traballos de campo. No momento de realizar a consulta, o *TLPGP* inclúe 155 obras que incorporan léxico galego. Dado o diferente carácter do *TLPGP* e dos corpus mencionados, as cifras ofrecidas no Apéndice non son comparables. No caso dos corpus proporcionamos o número de ocorrencias da forma buscada nos textos que os conforman e no caso do *TLPGP* o número de concellos en que se recolleu. Por outro lado, debe terse en conta que, se ben a presenza dunha forma neste tesouro pode considerarse en xeral unha proba do seu emprego no galego popular, a súa

⁶ No *TILG* ao redor do 44% das palabras pertencen a textos publicados antes da aprobación en 1982 da primeira normativa oficial do galego, mentres que no caso do *CORGA* esa porcentaxe baixa ao 2,9%.

ausencia non implica que non se utilice nesta variedade, posto que o vocabulario recollido nas obras do *TLPGP* adoita estar restrinxido a determinados ámbitos, sobre todo relacionados co mundo tradicional.

3.2.1. *Formas con diverxencias distintas ás de xénero e acentuación*

Inclúense neste grupo palabras como *plato* fronte ao estándar *prato*, *inconvincente*, *inconveniente* ou *inconvente* fronte a *inconveniente*, *asombrante* fronte a *asombroso* etc., que se etiquetan en *CORTEGAL* co código L_w_su. Quedan fóra as formas aceptadas no código normativo, aínda que non recomendadas, como *aportar*, co valor de ‘achegar’ ou *desfrutar*, que non se corrixen nin se etiquetan en *CORTEGAL*. Antes de analizar os datos rexistrados, cómpre delimitar máis claramente o grupo establecido, que unicamente integra desviacións de carácter formal e non inclúe os usos semanticamente inadecuados. Así, por exemplo, o verbo *facer* cando se emprega para expresar o tempo transcorrido («eu vía fai pouco estes programas») ou o substantivo *pobo* cando designa unha entidade de poboación («o pobo de Iago Aspas») considéranse desviacións do nivel semántico e non do léxico, na medida en que as palabras utilizadas son estándar, pero non así o emprego que delas se fai. Tampouco se consideran desviacións léxicas, senón discursivas, as formas cun rexistro impropio dos textos académicos, como *bestial* en «teñen unha bagaxe bestial».

Tamén se clasifican de maneira distinta, no nivel morfolóxico, aquelas desviacións do estándar que afectan a unha forma concreta do paradigma flexivo, pero non a toda a unidade léxica: así, por exemplo, *acude* (por *acode*) clasifícase como unha forma flexiva non estándar (M_v_su) e, do mesmo xeito, formas como *catalana*, *actuales*, *una* ou *esto* (por *catalá*, *actuais*, *unha* e *isto*) son consideradas no nivel morfolóxico, pois a selección inadecuada afecta a unha das formas do paradigma flexivo.

Cómpre tamén facer algunha aclaración sobre a delimitación con respecto ao nivel ortográfico. Optouse por considerar nesta última dimensión aquelas diverxencias que afectan a consoantes ou vogais e que non alteran ou poden non alterar a pronuncia, ou ben que representan unha pronuncia non estándar pero posible da palabra. Así, por exemplo, dado que no galego oral se tende a omitir en rexistros informais a pronuncia das oclusivas implosivas e mesmo a debilitarse en rexistros formais (Rogueira, 2010, p. 16) consideramos que hai unha desviación ortográfica, e non léxica, en *perfeto* (por *perfecto*) ou na ultracorrección *imperactivo* (por *imperativo*), e, daquela, tamén en *respeto* (por *respecto*) ou en *producción* (por *producción*), aínda que nestes

dous últimos casos parece clara a influencia dos cognados do español (e por tal motivo, estas dúas formas levan asignada no campo orixe a etiqueta O_sp, que permite localizalas de maneira independente). Por razóns similares, e dado que no galego oral popular son frecuentes as asimilacións e disimilacións vocálicas (Ferreiro, 1999, pp. 196, 200), consideráronse no nivel ortográfico formas como *devertíamos* (escrito *debertíamos*) (por *divertíamos*), *podiría* (por *podería*) ou *límete* (por *límite*), pero en coherencia codificamos igualmente nese nivel voces que tamén poden proceder directamente do español como *decir* (por *dicir*) ou *hipocresía* (por *hipocrisía*), que en calquera caso, para poderen ser localizadas separadamente, son tamén clasificadas cunha etiqueta de orixe (O_sp).

Tamén resulta difícil nuns poucos casos diferenciar entre os lapsus cálamis, codificados no nivel ortográfico, e as desviacións léxicas. En xeral, dado que tratamos con alumnado que estuda galego e en galego desde a educación primaria, tendemos a considerar formas anómalas para as que non cabe ningunha das explicacións sinaladas en §3.1 como lapsos (*alcalce* por *alcance*, *candidade* por *cantidad*, *adutos* por *adultos*, *sempren* por *sempre...*). Só nuns poucos casos en que non asignamos código de orixe, os comentados en §3.2.1.8 (*estrea*, por *estrela*, *reixear*, por *rexeitar*, *fugaxidade*, por *fugacidade*, e *lumen*, por *lume*), optamos por considerar a diverxencia ao nivel léxico.

Aínda con estas restricións no seu emprego, as formas con diverxencias con respecto ao estándar distintas ao xénero e á acentuación son, con moito, as desviacións máis numerosas do nivel léxico, como se pode comprobar na Táboa 1 presentada en §3.1. Hai un total de 2867 formas desviantes, 756 distintas, con 2887 etiquetas, posto que nalgúns casos se atribúen dous códigos L_w_su (por exemplo, en *perrito quente*, un código polo calco do español *perrito caliente* e outro polo mantemento do castelanismo *perrito*). Antes de analizar os exemplos, establecendo diferentes apartados en función da orixe da diverxencia con respecto ao estándar, ofreceremos algúns datos globais.

Na Táboa 5 proporcionamos a distribución das formas non estándar estudadas en función da súa orixe. Lémbrese que unha mesma forma pode recibir varios códigos deste tipo, de xeito que o número de códigos de orixe é superior a 2887. Como se indicou, en catro palabras non se asigna código deste tipo.

TÁBOA 5. Distribución das formas non estándar con diferenzas que van máis alá do xénero e da acentuación en función da súa orixe

Orixe	Código	Nº de códigos	Porcentaxe
Transferencias directas do castelán	L_sp	2135	71,10%
Transferencias do castelán con adaptación ou mestura co galego	L_spadapt	566	18,85%
Hipergaleguismos	L_hc	110	3,66%
Formas creadas por analogía a partir doutras voces galegas	L_anal	98	3,26%
Transferencias do galego popular ou do galego elaborado	L_gal	81	2,70%
Mestura ou confusión entre formas	L_for	7	0,23%
Estranxeirismos non adaptados	L_mix	6	0,20%
Total		3003	

Como se pode observar, e dado que estas formas supoñen máis do 95% das palabras e expresións léxicas non estándar, os resultados son moi similares aos datos globais ofrecidos na Táboa 2 (§3.1). Como alí, a orixe máis frecuente é a transferencia directa do español, seguida da transferencia con adaptación ou mestura co galego, sendo as restantes bastante minoritarias.

Igual que viamos para os datos globais, a maior parte dos casos que estamos analizando corresponden a ocorrencias únicas de formas non estándar: 422 das 756 formas distintas (un 55,82%). Ademais, 622, isto é, o 82,28% son palabras ou expresións con menos de 5 ocorrencias. Agora ben, en relación co que indicamos en §3.2, do nivel de ocorrencia en valores absolutos das distintas formas non podemos extraer conclusións sen máis sobre o grao de implantación das formas non estándar no galego do alumnado, ou sobre o maior ou menor grao de coñecemento das correspondentes voces normativas. Non podemos esquecer que o tipo e a temática dos textos condiciona o nivel de aparición das palabras. Así, por exemplo, hai 103 ocorrencias de *plato* (forma estándar *prato*) e 25 de *olvidar* (forma estándar *esquecer*), pero deste feito non podemos concluír a maior implantación de *plato* no galego, ou o menor nivel de coñecemento da forma estándar correspondente en relación con *esquecer*, pois a elevada presenza nos textos de *plato* responde a que un dos temas sobre os que debían escribir as/os estudantes está relacionado coa gastronomía: o concepto ‘prato’ está presente máis veces en *CORTEGAL* ca o concepto ‘esquecer’, o que determina que haxa máis ocorrencias en números absolutos de *plato* ca de *olvidar*, pero isto non implica necesariamente a súa maior presenza en termos relativos.

Así pois, para valorar adecuadamente a presenza das formas non estándar é preciso ter en conta tamén o nivel de presenza das correspondentes formas estándar. Limitándonos ás 134 formas non estándar con cinco ou máis ocorrencias, amosaremos a seguir unha táboa (Táboa 6) que mostra a súa distribución en franxas porcentuais, atendendo á súa presenza en relación co total de veces que aparece expresado o concepto. Así, por exemplo, dado que *plato* figura 103 veces e *prato* 408, *plato* ten un nivel de aparición do 20,2% en relación coas 511 ocorrencias do concepto, situándose na terceira franxa da Táboa 6. En cambio, dado que *olvidar* se rexistra 25 veces e *esquecer* 22, a porcentaxe de emprego do primeiro verbo para a expresión do concepto ‘esquecer’ é do 53,2%, situándose na sexta franxa.

No caso de que entre a lista das 134 formas con cinco ou máis ocorrencias atopemos algunha que teña algunha variante ou sinónimo non estándar no corpus, ao que lle corresponda o mesmo equivalente estándar, tales ocorrencias súmanse (con independencia do número de ocorrencias de que se trate): así, por exemplo, aos 8 resultados de *inconvinte*, súmase un de *inconvente* e outro de *inconveniente*. O mesmo aplica se existen varias realizacións posibles para unha forma estándar equivalente da forma non estándar rexistrada (por exemplo, súmanse as ocorrencias de *hostalaría* e *hostalería*). Na penúltima fila da Táboa 6 incluímos unha columna (100%) na que se rexistran aqueles casos, sempre con cinco ou máis ocorrencias, en que a voz estándar non se rexistra.

TÁBOA 6. Nivel de presenza das formas non estándar con cinco ou máis ocorrencias en relación coas correspondentes formas estándar

Franxa porcentual	Número de formas non estándar na franxa	Porcentaxe sobre o total	Porcentaxe acumulada
0,1%-9,9%	25	20,49%	20,49%
10%-19,9%	17	13,93%	34,42%
20%-29,9%	13	10,66%	45,08%
30%-39,9%	12	9,84%	54,92%
40%-49,9%	11	9,02%	63,94%
50%-59,9%	11	9,02%	72,96%
60%-69,9%	8	6,56%	79,52%
70%-79,9%	7	5,74%	85,26%
80%-89,9%	3	2,46%	87,72%
90%-99,9%	5	4,10%	91,82%
100%	10	8,20%	100%
Total	122		

Nótese que a cifra total do número de casos (122) é inferior á cifra total de formas con máis de cinco ocorrencias (134). Isto é debido a que ás veces se suman nun mesmo concepto dúas palabras con cinco ou máis resultados (por exemplo, *desarrollar* e *desenrolar* compúntanse conxuntamente) e tamén ao feito de que tivemos que deixar fóra unhas poucas palabras cunha ampla gama de equivalentes ou cun equivalente cun sentido demasiado amplo, o que dificultaba computar o nivel de presenza do concepto. Por este motivo, non consideramos *conlevar/conllevar* (34 ocorrencias), o pronome *elo* (10), *entorno* (7), *novidoso/novedoso* (17), *plantear/plantexar* (18) e *rato* (6).

Como se pode observar, existen 10 conceptos en que a forma estándar non se rexistra, pero en calquera caso, son maioritarios os exemplos en que a voz normativa é máis numerosa ca a forma non estándar. Tal e como se pode ver na última columna, isto sucede no 63,94% dos casos (fronte ao 36,06% das palabras en que a forma non estándar é a predominante). Na Táboa 7 ofrécese os 44 exemplos en que a forma non estándar presenta unha porcentaxe igual ou superior ao 50%. A columna ordenadora é a última, que indica a porcentaxe, sendo a segunda o número de ocorrencias das voces non estándar e a terceira a orde alfabética destas formas. Por suposto, aquí e nas seguintes táboas, cando a forma galega estándar ten máis sentidos ou máis contextos de uso ca a voz non estándar (por exemplo, *venda*), computamos unicamente os empregos en que hai equivalencia semántica e as voces son substituíbles⁷.

TÁBOA 7. Formas non estándar cun nivel de presenza igual ou superior ao das formas estándar correspondentes

Forma non estándar	Nº de casos	Forma estándar	Nº de casos	Porcentaxe de presenza da forma non estándar
<i>perxudicial</i>	18	<i>prexudicial</i>	0	100%
<i>afición</i> (14), <i>aficción</i> (1)	15	<i>afección</i>	0	100%
<i>hosteleiro</i> (12), <i>hostelero</i> (1)	13	<i>hostaleiro</i>	0	100%
<i>disfrute</i> (8), <i>goce</i> (2)	10	<i>gozo, disfrute</i>	0	100%
<i>postre</i>	9	<i>sobremesa</i>	0	100%
<i>tercermundista</i>	8	<i>terceiromundista</i>	0	100%

⁷ Aquí e nas restantes táboas do traballo, para a determinación dos equivalentes das transferencias do español acudimos ao *Diccionario castelán-galego da Real Academia Galega* (2005) e, no caso de non ter entrada a forma do español nesta obra, ao *Diccionario castelán-galego* da editorial Rinoceronte. En calquera caso, en ocasións cubrimos ou completamos a nómina de formas estándar con algunhas expresións que non figuran nos dicionarios citados e que propoñemos nós a partir da consulta doutras fontes.

TÁBOA 7. Continuación

<i>acarrear</i>	5	<i>carrear, carrar, carretar, carrexar</i>	0	100%
<i>cumpleanos</i>	5	<i>aniversario</i>	0	100%
<i>estropear</i>	5	<i>estragar, derramar</i>	0	100%
<i>inflúinte</i>	5	<i>influyente</i>	0	100%
<i>perxudicar</i>	33	<i>prexudicar</i>	1	97,06%
<i>invertir</i>	19	<i>investir</i>	1	95%
<i>apariencia</i>	16	<i>aparencia</i>	1	94,12%
<i>rexir</i>	18	<i>rexer</i> ⁸	2	90%
<i>chamativo (8), llamativo (1)</i>	9	<i>rechamante</i>	1	90%
<i>hostalería</i>	30	<i>hostalería (1), hostalaría (3)</i>	4	88,24%
<i>rentable (12), rentábel (2)</i>	14	<i>rendible</i>	3	82,35%
<i>promocionar</i>	8	<i>promover</i>	2	80%
<i>mezclar (9), misturar (2)</i>	11	<i>mesturar</i>	3	78,57%
<i>asequible</i>	18	<i>accesible</i>	5	78,26%
<i>tontería</i>	7	<i>parvada</i>	2	77,77%
<i>xunto con</i>	17	<i>xunto aº</i>	5	77,27%
<i>paladar (13), padar (1)</i>	14	<i>padal</i>	5	73,68%
<i>disfrutar</i>	92	<i>gozar (28), desfrutar (6)</i>	34	73,02%
<i>artiluxio</i>	5	<i>artefacto</i>	2	71,43%
<i>venta</i>	32	<i>venda</i>	16	66,66%
<i>basura</i>	19	<i>lixo</i>	10	65,52%
<i>búsqueda</i>	11	<i>busca</i>	6	64,71%
<i>inconvincente (8), inconveniente (1), inconvenciente (1)</i>	10	<i>inconveniente</i>	6	62,50%
<i>vacío</i>	10	<i>baleiro</i>	6	62,50%
<i>preciado</i>	5	<i>prezado</i>	3	62,50%
<i>soer (25), soler (3)</i>	28	<i>adoitar</i>	17	62,22%
<i>cercano</i>	9	<i>próximo</i> ¹⁰	6	60%
<i>mezcla (11), mistura (3)</i>	14	<i>mestura</i>	10	58,33%
<i>arreglar</i>	7	<i>arranxar</i>	5	58,33%

⁸ Non se computan nin en *rexir* nin en *rexer* nove casos do participio, que poderían corresponder a calquera das dúas conxugacións.

⁹ Non hai exemplos de *canda*.

¹⁰ Non hai exemplos de *achegado*.

TÁBOA 7. Continuación

<i>pilar</i>	14	<i>piar</i>	11	56%
<i>esclavitude</i>	5	<i>escravitude</i>	4	55,55%
<i>increíble</i>	6	<i>incrible</i>	5	54,54%
<i>coste</i> (29), <i>custe</i> (4), <i>costo</i> (3)	36	<i>custo</i>	31	53,73%
<i>olvidar</i>	25	<i>esquecer</i>	22	53,19%
<i>receta</i>	38	<i>receita</i>	35	52,05%
<i>costar</i>	16	<i>custar</i>	16	50%
<i>reflexar</i> (12), <i>reflejar</i> (1)	13	<i>reflectir</i>	13	50%
<i>grado</i>	8	<i>grao</i>	8	50%

Como se pode observar, entre as formas non estándar atopamos sobre todo voces coincidentes co español, pero tamén palabras adaptadas (como *reflexar*) ou mesmo creadas seguindo mecanismos produtivos do galego (como *inconvinte*). Por outro lado, de maneira xeral, tal e como se pode comprobar na Táboa 38 do Apéndice, as formas non estándar da Táboa 7 teñen certa presenza nos corpus estudados (*CORGA* e *TILG*), con máis de 20 ocorrencias cando menos nun deles. Deixando a un lado as variantes consideradas a maiores e que teñen menos de cinco exemplos en *CORTEGAL*, as únicas excepcións son as palabras *esclavitude*, *cumpleanos* e *tercermundista*, que en todo caso se rexistran cando menos unha vez nalgún dos dous corpus. Ademais, 20 das 58 variantes da táboa (36,21%) aparecen recollidas no *TLP-GP*, pero, tal e como xa indicamos, e dadas as limitacións temáticas do vocabulario que adoita figurar nas obras incluídas nesta base de datos, a ausencia nela dunha forma non implica que non se utilice no galego popular.

A seguir, analizaremos con máis detalle as formas con diverxencias con respecto ao estándar que van máis alá do xénero ou da acentuación, establecendo diferentes apartados en función da súa orixe.

3.2.1.1. Transferencias directas do español

Analizamos neste grupo aquelas formas non estándar coincidentes co español. Debe sinalarse que as formas cuxo tema coincide co castelán e simplemente se adecúan ás características flexivas do galego (*saliu*, *rincóns*, *pañais*...) tamén son incluídas aquí.

Son un total de 2135 códigos (o 71,10% dos códigos de orixe asignados ás formas con diferenzas que van máis alá do xénero e da acentuación), repartidos entre 508 formas diferentes, o que supón unha media de 4,20 ocorrencias por forma, cunha desviación estándar de 8,79 e un rango que se

sitúa entre os 103 exemplos de *plato* e 263 casos cunha única ocorrencia. Incluímos dentro do cómputo aquelas formas que se integran en calcos con tradución parcial (*vid.* §3.2.1.2), como por exemplo *perrito* en *perrito quente* ou *cumple-* en *cumpleanos*.

Dado o elevado número de formas deste grupo, non ofreceremos a lista-xe de todas elas, que se poden consultar no buscador de *CORTEGAL* seleccionando L_w_su en «Tipo de problema» e L_sp en «Orixe do problema». En calquera caso, poden verse algunhas delas na Táboa 7, pois, tal e como se pode comprobar nesta, a maior parte das formas non estándar cunha porcentaxe igual superior ao 50% de realización do concepto son voces coincidentes co español.

No conxunto de transferencias directas do español encóntranse voces ben testemuñadas no galego popular e no galego elaborado, como por exemplo *olvidar*, *plato*, *salir* ou *según* (todas elas teñen entrada no *TLPGP* e varios rexistros no *CORGA* e no *TILG*), as cales poderían ser formas integrantes do lexicón mental das/dos estudantes que as escriben (préstamos para Gómez Molina, 1999-2000, pp. 323-324), mentres que outras son formas que probablemente respondan a transferencias espontáneas (interferencias para Gómez Molina, 1999-2000, p. 323), isto é, «el resultado de actos idiolectales», os cales poden explicarse «por descuido, desconocimiento, escasa competencia bilingüe etc.» ou ben poden responder «a alguna estrategia utilizada por el hablante» (Gómez Molina, 1999-2000, p. 323). Dado que a/o estudante sabe que as desviacións do galego estándar serán sancionadas, o que exclúe desvíos conscientes con funcionalidade pragmática, dentro deste grupo de interferencias incluíríanse tanto formas para as que a/o estudante formula hipóteses erradas de coincidencia en galego e castelán (dada a proximidade de ambas as linguas e a elevada porcentaxe de léxico común), como lapsos bilingües («bilingual slips» para Matras, 2009, p. 19), despsites ou lapsos na selección da lingua á hora de escribir os textos, sobre todo por parte de persoas que teñen o castelán como lingua habitual (Paradis, 2004, p. 188 fala de «dynamic interferences»). Así, por exemplo, un/unha estudante escribe «Estes programas animan a la gente a cocinar», con varios elementos seguidos en español (*la*, *gente*, *cocinar*) entre os que mesmo se inclúe o artigo definido. Outras formas non recollidas no galego popular nin no galego elaborado, pero si en *CORTEGAL* son os posesivos *mi* ou *su* (no mesmo texto, figura *seu*), o numeral *dos*, verbos como *tener*, *poner*, *hacer* ou *venir*, substantivos como *cosa* ou *año* (a mesma persoa escribe no mesmo texto *ano*) etc. En calquera caso, son, pola súa propia definición, casos

illados, cunha porcentaxe que, cando menos nos exemplos citados, non chega nunca ao 1% de ocorrencias no total do concepto.

Tal e como se pode ver nos exemplos anteriores, as formas transferidas desde o castelán pertencen a distintas clases de palabras, incluídas palabras gramaticais, pero son as transferencias de substantivos as máis frecuentes, coherentemente co indicado na Táboa 4 para o conxunto de formas non estándar. Lembremos que a transferencia directa do español é a causa máis frecuente dos desvíos da norma en *CORTEGAL*, de modo que é esperable que non haxa grandes diferenzas entre os resultados globais e os correspondentes a esta orixe. De feito, na Táboa 8, en que se recolle a distribución das formas por categorías de palabras, pode comprobarse como os resultados son bastante similares aos atopados na Táboa 4.

TÁBOA 8. Distribución por categorías gramaticais das transferencias directas do español (con diferenzas que van máis alá do xénero e a acentuación)

Categoría gramatical	Porcentaxe
Substantivos	46,16%
Verbos	26,10%
Adxectivos	10,60%
Adverbios	6,12%
Conxuncións	4,29%
Pronomes e determinantes	2,40%
Preposicións	2,26%
Expresións complexas	1,88%
Interxeccións	0,14%
Numerais	0,05%

3.2.1.2. Transferencias do español con adaptación ou mestura co galego

Agrupamos aquí un conxunto de 139 formas diferentes, cun total de 566 códigos no corpus e unha media de 4,07 ocorrencias por forma. Pertencen a este grupo o 18,85% das formas con diferenzas que van máis alá do xénero e da acentuación que foron clasificadas en canto á orixe da desviación. Trátase de palabras ou expresións con algún tipo de adaptación ao galego desde unha voz ou expresión do castelán ou con algunha mestura, menos sistematizable, entre unha voz do galego e o seu cognado español. Debe terse en conta, en calquera caso, que se trata dunha adaptación que vai máis alá da flexión, pois, tal e como indicamos en §3.2.1.1, as formas cuxo tema coincide co castelán e

simplemente se axeitan á flexión do galego (*saliu, rincóns...*) son consideradas transferencias directas do español.

Un total de 12 das formas con máis de 5 ocorrencias incluídas na Táboa 7, táboa en que se recollían os casos en que as formas non estándar tiñan unha porcentaxe de aparición igual ou superior ás correspondentes formas estándar, pertencen a este grupo (*artiluxio, chamativo, cumpleaños, esclavitude, hosteleiro, influínte, perxudicar, perxudicial, reflexar rexir, soer e xunto con*).

Dentro do grupo de formas estudadas, incluímos, por unha banda, aquelas palabras ou expresións que reciben o nome de calcos estruturais ou calcos do esquema (Gómez Capuz, 1998, p. 63), ás veces simplemente calcos (Rodríguez Río, 2003, p. 422): neles, «partindo de dous ou máis significantes preexistentes na lingua receptora créase unha combinación inédita copiando as pautas polas que se formou noutra lingua» (Rodríguez Río, 2003, pp. 422-423).

No grupo destes calcos estruturais, encontramos por un lado algúns exemplos en que o elemento que se calca é unha expresión fixa constituída por varias palabras. Na primeira columna da Táboa 9 presentamos as expresións calcadas, na segunda columna os exemplos rexistrados, con indicación entre parénteses do número de ocorrencias, mentres que na terceira columna se indica o número de veces que encontramos en *CORTEGAL* a(s) forma(s) estándar(es) equivalentes (en contextos en que serían intercambiabes).

TÁBOA 9. Calcos estruturais de expresións pluriverbais fixas do español

Expresión fixa do español	Calco estrutural	Forma estándar
<i>sin embargo</i>	<i>sen embargo</i> (52)	<i>porén</i> (63), <i>non obstante</i> (43), <i>con todo</i> (7), <i>así e todo</i> (5), <i>no entanto</i> (1), <i>así a todo</i> (0)
<i>junto con</i>	<i>xunto con</i> (17)	<i>xunto a</i> (5), <i>canda</i> (0)
<i>a la larga</i>	<i>á longa</i> (2)	<i>co tempo</i> (15), <i>a longo prazo</i> (2)
<i>hoy día</i>	<i>hoxe día</i> (2)	<i>hoxe en día</i> (365)
<i>una pescadilla que se muerde la cola</i>	(unha) pescadilla que se morde / mórdese a cola (2)	(unha) <i>carioca co rabo na boca</i> (0), (un) <i>peixe que morde a cola</i> (0)
<i>la pescadilla que se muerde la cola</i>	(a) <i>pescada que se morde a cola</i> (1)	(a) <i>carioca co rabo na boca</i> (0), (un) <i>peixe que morde a cola</i> (0)
<i>comerse el mundo</i>	<i>comerse</i> o mundo (1)	<i>conquistar</i> (dominar...) o mundo (0)
<i>dárselas</i>	<i>dárseas</i> (1)	<i>ir de</i> (0)

TÁBOA 9. Continuación

<i>de campeonato</i>	<i>de campionato</i> (1)	<i>de antoloxía</i> (o), <i>de categoría</i> (o), <i>de manda chover na Habana</i> (o), <i>de manda truco</i> (o), <i>de manda narices</i> (o), <i>de marca maior</i> (o), <i>de moito pendello</i> (o),
<i>perrito caliente</i>	perrito quente (1)	<i>bocadillo de salchicha</i> (o), <i>bolo con salchicha</i> (o)
<i>picar el gusanillo</i>	<i>picar o gusaniño</i> (1)	<i>entrar o bicho</i> , (o) <i>meterse o bicho no corpo</i> (o), <i>picar a curiosidade</i> (o)
<i>recién nacido</i>	recén nacido (1)	(<i>nen</i>) <i>acabado de nacer</i> (o), <i>neonato</i> (o)
<i>repartirse el pastel</i>	<i>repartirse o pastel</i> (1)	<i>distribuír entre si</i> (o)
<i>tira y afloja</i>	<i>tira e afloxa</i> (1)	<i>axusta e compón</i> (o), <i>tirapuxa</i> (o)

Ademais da hibridación implícita no calco (estrutura do español e materiais do galego), na maior parte dos exemplos ofrecidos percíbese un segundo nivel de influencia do castelán, destacado en letra grosa na segunda columna da táboa. Por un lado, debe notarse que se inclúen neste grupo varios exemplos de expresións que non se adecúan ás pautas sintácticas do galego, pola presenza dun pronome *se* en contextos ou con valores que resultan agramaticais nesta lingua e que se herdán directamente da estrutura do español: *unha pescadilla que se morde a cola / a pescada que se morde a cola, comerse o mundo, dárseas*¹¹ e *repartirse o pastel*. Por outra banda, tamén cómpre destacar que nalgúns casos a «tradución» non é completa, porque algún dos elementos do calco se mantén na forma do español, como no caso de *pescadilla* e *perrito*, dúas formas que ademais portan senllos sufixos diminutivos característicos do castelán e alleos á morfoloxía do galego. Probablemente este último feito explica que nun dos textos atopemos unha versión lixeiramente máis libre da expresión, ao traducirse *pescadilla* por *pescada*, eliminando o morfema apreciativo¹². Así mesmo, en *picar o gusaniño*, o substantivo *gusaniño* mantén a raíz do español, pero o diminutivo *-illo* é substituído por *-iño*, en *tira e afloxa*, o calco bota man dunha adaptación fónica do castelán

¹¹ A combinación do pronome *se* e do clítico de obxecto directo de terceira persoa é agramatical en galego. Concretamente, a forma escrita pola/polo estudante é *dansenas* («estes cocifeiros premiados dansenas de superiores»).

¹² *Pescada* é o equivalente galego do español *merluza*, e a *pescadilla* é unha forma lexicalizada do castelán que designa unha *merluza* pequena. O equivalente que adoita ofrecerse para *pescadilla* en galego é *carioca* (vid. por exemplo, TERGAL http://bernal.cirp.gal/ords/f?p=121:8:::P8_IDT:59531).

aflojar, frecuente na fala (vid. o *TLPGP*), e o mesmo sucede no caso de *recén nacido*, onde a forma *recién* é substituída por *recén*.

Outro grupo de calcos é o constituído por compostos léxicos e sintagmas unverbalizados, tal e como se recolle na Táboa 10.

TÁBOA 10. Calcos estruturais de compostos léxicos e sintagmas unverbalizados do español

Composto léxico / sintagma unverbalizado	Calco estrutural	Forma estándar
<i>cualquiera</i>	<i>calqueira</i> (25)	<i>calquera</i> (119)
<i>cumpleaños</i>	<i>cumpleanos</i> (5)	<i>aniversario</i> (0)
<i>sobremanera</i>	<i>sobremaneira</i> (2)	<i>moitísimo</i> (10), <i>enormemente</i> (7), <i>excesivamente</i> (2), <i>extraordinariamente</i> (0), <i>fortemente</i> (1), <i>extremadamente</i> (0), <i>profundamente</i> (0), <i>en gran maneira</i> (0)
<i>engañabobos</i> (0)	<i>enganabobos</i> (1)	<i>engano</i> (2)

Os catros exemplos poden explicarse como traducións dos compoñentes que conforman cada unidade, aínda que no caso de *calqueira* cabe tamén outra explicación consistente no cambio da terminación *-era* por *-eira* a partir de *calquera* (vid. 3.2.1.3). Unha vez máis, a tradución pode ser «imperfecta» desde a perspectiva do estándar, pois tamén neste grupo encontramos un segundo nivel de influencia do español en *cumpleanos*, onde o primeiro compoñente aparece na forma coincidente co castelán (*cumplir* vs. a forma estándar *cumprir*). O verbo *cumplir*, como a forma *cumpleanos*, é frecuente na fala popular (vid. *TLPGP*) e aparece recollido como forma independente nove veces en *CORTEGAL*.

En terceiro lugar, na Táboa 11 ofrecemos voces (na súa maior parte derivados mediante afixación) que copian as pautas formativas de palabras do español, traducindo a raíz e mantendo os sufixos e/ou os prefixos, os cales, en todos os exemplos rexistrados, son comúns a ambas as dúas linguas. Para *conlevar* e *novidoso* non ofrecemos cómputos da forma estándar polas razóns expostas en §3.2.1:

TÁBOA 11. Calcos estruturais de palabras derivadas e outras palabras complexas

Forma do español	Raíz ou base do galego	Calco estrutural	Forma estándar
<i>conllevar</i>	<i>levar</i>	<i>conlevar</i> (33)	<i>implicar, supoñer, supor, comportar</i> (--)
<i>novedoso</i>	<i>novidade</i>	<i>novidoso</i> (11)	<i>novo</i> (--)
<i>llamativo</i>	<i>chamar</i>	<i>chamativo</i> (8)	<i>rechamante</i> (1)
<i>llenar</i>	<i>cheo</i>	<i>chear</i> (7)	<i>encher</i> (23)
<i>coste</i>	<i>custar</i>	<i>custe</i> (4)	<i>custo</i> (31)
<i>lejano</i>	<i>lonxe</i>	<i>lonxano</i> (4)	<i>afastado</i> (3), <i>remoto</i> (1), <i>distante</i> (0)
<i>alejar</i>	<i>lonxe</i>	<i>aloxar</i> (1)	<i>afastar</i> (5), <i>apartar</i> (5), <i>arredar</i> (0), <i>alongar</i> (0)
<i>temprano</i> (adx.)	<i>cedo</i> (adv.)	<i>cedo</i> (1) (adx.)	<i>temperán</i> (1)
<i>dañino</i>	<i>dano</i>	<i>danino</i> (1)	<i>daniño</i> (1)
<i>dolencia</i>	<i>doer</i>	<i>doencia</i> (1)	<i>doenza</i> (1)
<i>hornear</i>	<i>forno</i>	<i>formear</i> (1)	<i>enformar</i> (1)
<i>rellenar</i>	<i>cheo</i>	<i>rechear</i> (1)	<i>reencher</i> (0)
<i>releño</i>	<i>cheo</i>	<i>recheo</i> (adx.) (1)	<i>gordo</i> (2), <i>gordecho</i> (0), <i>groso</i> (0)
<i>trueque</i>	<i>trocar</i>	<i>troque</i> (1)	<i>troco</i> (2)

Cómpre comentar, polo seu carácter particular, o caso de *cedo* empregado como adxectivo («Normalmente, a ceda idade dos 15-16 anos, os fillos entran na coñecida idade «do pavo»»). Cremos que moi probablemente o feito de que en español a forma *temprano*, equivalente do adverbio galego *cedo*, sexa tamén adxectivo é o que explica que a voz galega se reconverte a estoutra categoría, facéndose variable mediante a conversión da vogal final en afixo flexivo, como sucede na voz do español. Deste cambio categorial de *cedo* temos outros testemuños en corpus galegos: no *TILG*, con 19 exemplos, e no *CORGA*, con 17 casos.

Con respecto precisamente á presenza dos calcos noutras fontes, e tal e como se pode comprobar na Táboa 39 do Apéndice, só dous deles se rexistran no galego popular do *TLPGP*, en ambos os casos cunha soa entrada na base de datos: *rechear* e *calqueira*¹³. *Alonxar*, *calqueira*, *chamativo*, *conllevar*, *custe*, *doencia*, *lonxano*, *novidoso*, *recén nacido*, *rechear*, *recheo*, *sen*

¹³ Esta forma rexístrase na obra de González González (2002) sobre o galego do concello Cangas do Morrazo.

embargo, sobremaneira, troque e xunto con son expresións ben representadas no *TILG* e no *CORGA*, con cando menos 20 ocorrencias nos dous corpus ou nalgún deles, mentres que só catro expresións, catro calcos de expresións fixas do español, están totalmente ausentes das tres fontes: *perrito quente, (unha) pescadiña que se morde a cola, picar o gusaniño e repartirse o pastel*.

Ao lado dos calcos, encontramos numerosos exemplos de formas importadas do español (ou do galego popular, que contén numerosos castelanismos léxicos, Dubert, 2005) que experimentan un proceso de adaptación ao galego (*palabras adaptadas / integradas / acomodadas / asimiladas*, Haspelmath, 2009, p. 42). A adaptación consiste na aplicación dunha serie de modificacións, que afectan a sons ou a secuencias de sons, e que obedecen a padróns de equivalencia ou correlacións recorrentes entre galego e castelán, derivadas do diferente devir histórico de cada lingua. Algúns dos exemplos ofrecidos nas táboas que se presentan a seguir poderían tamén considerarse calcos parciais, na medida en que se produce a adaptación de elementos sufixais, mantendo a raíz do español (por exemplo, *aislamiento*). Moito máis raramente sucede ao revés, pois o único caso rexistrado é o de *productividad*, en que se mantén o sufixo do español (*-idad*) e a raíz coincide coa forma galega (*productivo*).

Na Táboa 12 ofrécense aquelas modificacións que afectan a consoantes:

TÁBOA 12. Adaptacións que afectan a elementos consonánticos

Padrón		Ocorrencias			Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma adaptada	Forma estándar	Forma non estándar		
<g[e,i]>, <j>	<x>	<i>perjudicar, joven, perjudicial...</i>	<i>prexudicar (1), mozo (32) / novo (21), prexudicial (0)...</i>	<i>perxudicar (33), xoven (26), perxudicial (18)...</i>	179	38
<n> intervoc.	<ñ>	<i>entretenimiento, tener, mantener...</i>	<i>entretemento (17), ter (401), manter (57)...</i>	<i>entreteñemento (2), teñer (2), manteñer (1)...</i>	6	4
<n>, <l> intervoc.	--	<i>soler, entrenar</i>	<i>adoitar (17), adestrar (0)</i>	<i>soer (25), entrear (1)</i>	26	2
<ll>	<l>	<i>ello, desarrollar</i>	<i>isto / iso... (--), desenvolver (65)</i>	<i>elo (10), desarrolar (1)</i>	11	2

TÁBOA 12. Continuación

<ll>	<i>	<i>estrella</i>	<i>estrela</i> (39)	<i>estreia</i> (1)	1	1
<pl>	<pr>	<i>empleo</i>	<i>emprego</i> (19)	<i>emprego</i> (1)	1	1
<uct>	<ut>	<i>productividad</i>	<i>produtividade</i> (198)	<i>produtividade</i> (1)	1	1

O exemplo máis evidente e recorrente de adaptacións deste tipo encontrámo-lo nas palabras que en castelán teñen o son [x] (representado mediante <g> ou <j>) e que muda para o son [ʃ], representado mediante <x>. Nótese que o son [x] non forma parte do sistema consonántico do galego e que a correspondencia castelán [x] / galego [ʃ] é moi habitual entre cognados (*jamón* / *xamón*, *ágil* / *áxil*...). O total de exemplos rexistrados pode consultarse na Táboa 40 do Apéndice. Agás no caso de *surximento* (vs. *surgimiento*) e *inxenieiro* (vs. *ingeniero*), a única diferenza existente entre a voz recollida no corpus e a palabra do español é a presenza de <x> en vez de <g> ou <j>.

Nos restantes casos prodúcese a substitución doutros sons consonánticos: en *teñer*, *manteñer* e *sosteñer*, infinitivos proporcionados por catro estudantes diferentes, o <n> intervocálico pasa a <ñ>, o cal pode explicarse pola correlación castelán <n> / galego <ñ> que encontramos en varias palabras como *camino* / *camiño* ou *ganar* / *gañar*. Con todo, tamén é posible apelar a unha formación analóxica a partir doutras formas do paradigma do verbo, como o presente de subxuntivo (*teña*, *teñas*...) e a P1 e a P6 do presente de indicativo (*teño*, *teñen*) ou, máis probablemente, á influencia destas formas na adaptación desde o castelán. Outro exemplo similar encontrámo-lo en *entreteneñemento* (proporcionada por dous/dúas estudantes diferentes das/dos que escribiron os infinitivos anteriores), que resultaría dun proceso derivativo a partir da forma *entreteneñer*.

En *emprego*, de súa vez, prodúcese a substitución da lateral do grupo consonántico <pl> por unha rótica, seguindo o modelo de *plata* / *prata*, *plaza* / *praza*..., mentres que en *desarrollar* e no pronome *elo* o <ll> pasa a <l> (como ocorre en cognados como *villa* / *vila* ou *ella* / *ela*). Nótese que *elo* é unha palabra con certa presenza en discursos formais (Fernández Salgado, 2004, s.v.) en cuxa forma tamén pode influír o correspondente pronome feminino tónico do paradigma pronominal *ela(s)* (vid. §3.2.1.4). Con respecto a *desarrollar*, podería tamén considerarse un calco en que a raíz *rollo* se traduce por *rolo*, se ben é certo que a identificación deste substantivo como compoñente do derivado non parece ser automática. No caso de *estreia*, o que atopamos é a substitución de <ll> por <i>. A correspondencia <ll> / <i>

non é habitual entre cognados, pero si encontramos varios casos de correspondencia entre <y> e <i> (*rayo / raio, apoiar / apoiar* etc.) e non pode esquecerse que <ll> e <y> en posición intervocálica representan o mesmo son no castelán de Galicia, posto que este é esencialmente yeísta (Dubert, 2013, p. 39).

Noutro grupo de formas encontramos uns poucos casos de adaptacións que consisten na eliminación de consoantes, concretamente de <n> ou <l> intervocálicos, seguindo unha vez máis correlacións entre cognados (*cena / cea, luna / lúa; muela / moa, volar / voar...*). Os exemplos rexistrados son *entrear*, a partir de *entrenar*, e *soer*, a partir de *soler*. Nótese, en calquera caso, que a frecuente presenza en CORTEGAL da forma *soer* pode tamén ter a súa orixe na revitalización que experimentou o arcaísmo galego *soer* en rexistros formais e literarios (Fernández Salgado, 2004, s.v. *soler*) (§3.2.1.5).

Finalmente, encontramos o caso xa comentado, en que se produce a adaptación da raíz, coa eliminación do <c> en posición implosiva (seguindo o modelo de *producto / produto* ou *acueducto / acueduto*), manténdose o elemento sufixal do español (-*idad*). Nótese que a/o estudante, no mesmo texto en que escribe *produtividade*, escribe tamén as formas *productividade* e *produto*.

En dous casos o que atopamos é unha adaptación da acentuación gráfica da palabra, tal e como se ve na Táboa 13.

TÁBOA 13. Adaptacións que afectan á acentuación

Padrón		Ocorrencias			Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma adaptada	Forma estándar	Forma non estándar		
<ui>	<uí>	<i>disminuir, huir</i>	<i>diminuír</i> (28), <i>fluxir</i> (3)	<i>disminuír</i> (1), <i>huír</i> (1)	2	2

Os dous exemplos atopados son *disminuír* e *huír*, en que as formas coincidentes co castelán se acentúan de acordo coas regras do galego (cf. cast. *atribuir, constituir* / gal. *atribuír, constituír*). En galego, as vogais <i> e <u> tónicas acentúanse graficamente «cando van inmediatamente antes ou despois dunha vogal átona, para indicar que ambas as vogais pertencen a sílabas distintas e que, polo tanto, non forman ditongo» (*Normas ortográficas e morfolóxicas do idioma galego*, Real Academia Galega, 2012, p. 25).

Noutro grupo de exemplos, a adaptación afecta ao vocalismo, concretamente a ditongos. Son os que se recollen na Táboa 14:

TÁBOA 14. Adaptacións que afectan a ditongos

Padrón		Ocorrencias			Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma adaptada	Forma estándar	Forma non estándar		
<ie>	<e>	<i>cualquier, mientras, pimiento...</i>	<i>calquera</i> (119), <i>mentres</i> (74), <i>pemento</i> (2)...	<i>calquer</i> (29), <i>mentras</i> (8), <i>pimento</i> (2)...	49	11
<cu>	<ca>	<i>cualquier</i>	<i>calquera</i> (119)	<i>calquer</i> (29)	29	1
<ue>	<o>	<i>trueque</i>	<i>troco</i> (2)	<i>troque</i> (1)	1	1

En todos os casos, un ditongo convértese nunha vogal. Por un lado, encontramos exemplos en que a secuencia <cu> pasa a <ca> (*calquer*), seguindo o modelo de *cuatro* / *catro* ou *cuando* / *cando*, e <ue> a <o> (*troque*), seguindo o modelo de *puerta* / *porta*, *zueco* / *zoco*...

Por outro lado, encontramos varios casos en que a secuencia <je> pasa a <e> (pode verse a listaxe completa na Táboa 40 do Apéndice). Esta adaptación vén dada pola recorrencia da equivalencia <ie> / <e> entre cognados: *piedra* / *pedra*, *fiera* / *fera*... Boa parte dos exemplos son palabras acabadas no sufixo *-mento* (*aislamiento*, *entendimiento*, *razonamiento*, *reconocimiento*, *rendimiento*, *replanteamiento*), de xeito que en tales casos poderíamos falar tamén de calcos en que se traduce o elemento afixal (*-miento* pasa a *-mento* antes ca <ie> se adapta a <e>) ou mesmo de creacións galegas sobre o tema do español. Unha interpretación como calco parcial tamén podería ser válida para a forma *hincapé*, no caso de que a/o estudante identificase esta voz como un composto cuxo segundo elemento é o substantivo *pie*, que pasaría a *pé*.

Os restantes exemplos caracterízanse porque neles se produce a adaptación de terminacións que son en boa parte dos casos, aínda que non sempre, sufixos. Son os que se recollen na Táboa 15.

TÁBOA 15. Adaptacións que afectan a terminacións

Padrón		Ocorrencias			Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma adaptada	Forma estándar	Forma non estándar		
<-ero, a>	<-eiro, a>	<i>hostelero, financiero, ingeniero...</i>	<i>hostaleiro</i> (0), <i>financeiro</i> (0), <i>enxeñeiro</i> (1)	<i>hosteleiro</i> (12), <i>financeiro</i> (4), <i>inxenheiro</i> (3)...	28	10
<-d>	<-de>	<i>esclavitud, novedad, pubertad...</i>	<i>escravitude</i> (4), <i>novidade</i> (17), <i>puberdade</i> (0)...	<i>esclavitude</i> (5), <i>novidade</i> (4), <i>pubertade</i> (2)...	20	11
<-ear>	<-exar>	<i>plantear</i>	<i>formular / presentar / considerar...</i> (--)	<i>plantexar</i> (16)	16	1
<-ano, -ana>	<-án, á>	<i>cotidiana, cotidiano</i>	<i>cotía</i> (20), <i>cotían</i> (2)	<i>cotidiá</i> (2), <i>cotidián</i> (2),	4	1
<-imiento>	<-emento>	<i>desvanecimiento, entretenimiento</i>	<i>esvaeacemento</i> (0), <i>entretemento</i> (17)	<i>desvanecemento</i> (1), <i>entretenemento</i> (1)	2	2
<-ble>	<-bel>	<i>rentable</i>	<i>rendible</i> (3)	<i>rentábel</i> (2)	2	1
<-ería>	<-aría>	<i>ganadería</i>	<i>gandaría</i> (5), <i>gandería</i> (1)	<i>ganadaría</i> (1)	1	1
<-illo>	<-iño>	<i>gusanillo</i> (<i>picar el gusanillo</i>)	<i>verme</i> (0), <i>vermiño</i> (0)	<i>gusaniño</i> (1)	1	1
<-dumbre>	<-dume>	<i>incertidumbre</i>	<i>incerteza</i> (0)	<i>incertidume</i> (1)	1	1
<-en>	<-e>	<i>resumen</i>	<i>resumo</i> (33)	<i>resume</i> (1)	1	1

Como xa indicamos, no caso de que as terminacións sexan sufixos, podería falarse tamén de calcos parciais, en que se mantén a raíz e se traduce o elemento sufixal. En primeiro lugar, encontramos varios exemplos de adaptación da terminación *-ero*, *-era* en *-eiro*, *-eira*, seguindo o modelo de *acero* / *aceiro* ou *caldera* / *caldeira*. A nómina completa pode verse na Táboa 40 do Apéndice (as voces *terneiro* e *terneira* tamén poden ser formas dialectais de orixe galega, *vid.* §3.2.1.5). Outro grupo numeroso está formado por palabras rematadas en <d> en español, maioritariamente acabadas no sufixo *-dad*, que toman un <e> final. Nótese que non hai en galego palabras patrimoniais rematadas nesa consoante e que o padrón que se repite nos cognados é o de

español *-d* / galego *-de*: *pared* / *parede*, *claridad* / *claridade*, *abad* / *abade*. A listaxe completa de formas adaptadas segundo este padrón pode verse na mesma Táboa 40 do Apéndice.

Por outro lado, encontramos dous exemplos de substantivos derivados dun verbo da segunda conxugación rematados no alomorfo *-emento* (*desvanecemento* e *entretenemento*), en que se pode postular a adaptación de <-imiento> (*desvanecimiento* e *entretenimiento*) en <-emento> seguindo o padrón de *nacimiento* / *nacemento* ou *movimiento* / *movemento* (cf. as formas *entendimiento*, *reconocimiento* e *rendimiento*, presentadas na Táboa 14, onde a vogal temática verbal é a do español). As formas rexistradas poderían ser tamén o resultado da creación directa en galego sobre as transferencias do español *desvanecer* e *entretener*.

Casos cun único exemplo distinto encontrámoslos en *cotidián* / *cotidiá* (aínda que desta volta temos unha forma diferente para masculino e para feminino), que resulta de adaptar a terminación *-ano*, *-ana* de *cotidiano*, *-ana* a *-án*, *-á* seguindo correspondencias como *aldeano*, *-na* / *aldeán*, *-á* ou *artesano*, *-na* / *artesán*, *-sá*; en *incertidume*, que provén da adaptación de *incertidumbre* seguindo o modelo de *pesadumbre* / *pesadume* ou *mansedumbre* / *mansedume*; en *plantexar*, unha palabra moi frecuente en discursos formais, creada a partir de *plantear* e seguindo o modelo de *cacarear* / *cacarexar*, *pestanear* / *pestanexar*...; en *resume*, en que a terminación *-en* de *resumen* perde o *-n* e pasa a *-e*, como tamén ocorre en *virgen* / *virxe* ou en *examen* / *exame*, así como en *gusaniño*, na expresión *picar o gusaniño*, xa mencionada, onde o diminutivo do castelán *-illo* é substituído polo apreciativo galego *-iño*. Cómpre sinalar ademais dous casos en que se modifican sufixos coincidentes en galego e castelán (*-ble* e *-ería*) para substituílos por variantes formais galegas de idéntico significado e diferenciais con respecto ao español (respectivamente, *-bel* e *-aría*): *rentábel* e *ganadaría*.

De acordo cos datos recollidos na Táboa 40 do Apéndice, só unhas poucas formas adaptadas aparecen rexistradas no *TLPGP*, entre as que predominan aquelas en que se produce o paso da fricativa velar á postalveolar, como ocorre con *xoven*, *xuez*, *parexa*, *antoxo*... No que respecta á presenza nos corpus, 29 das formas estudadas (un 31,52%) teñen nula ou escaza presenza neles (menos de 20 exemplos en ambos), de xeito que a maioría das formas están relativamente ben representadas nos textos. Só 11 das 94 palabras están ausentes de todas as fontes estudadas: *aconsexable*, *autoperxudicar*, *cocineiro*, *dineiro*, *estreia*, *ganadaría*, *hincapé*, *huír*, *produtividade*, *teñer* e *trampantoxo*.

Ao lado das formas citadas ata o de agora encontramos outras que combinan elementos das voces galegas e castelás, pero sen que sexa posible atopar un padrón regular. Así por exemplo, na forma *inxeniería*, rexistrada nunha ocasión en *CORTEGAL*, a primeira sílaba (<in->) coincide coa primeira sílaba da voz do español mentres que o resto da palabra o fai coa voz galega estándar. Os exemplos rexistrados, cun máximo de tres ocorrencias por forma (no caso de *costo*¹⁴ e *ventaxe*), son as que se rexistran na Táboa 16.

TÁBOA 16. Combinacións de galego e castelán sen padrón regular

Castelán estándar	Forma galega estándar	Forma galega non estándar
<i>ventaja</i>	<i>vantaxe</i> (19)	<i>ventaxe</i> (3) e <i>vantaxa</i> (2)
<i>coste</i>	<i>custo</i> (31)	<i>costo</i> (3)
<i>desventaja</i>	<i>desvantaxe</i> (3)	<i>desvantaxa</i> (2) e <i>desventaxe</i> (1)
<i>objeto</i>	<i>obxecto</i> (62)	<i>objecto</i> (2)
<i>afición</i>	<i>afección</i> (0)	<i>aficción</i> (1)
<i>aficionar</i>	<i>afeccionar</i> (2)	<i>aficcionar</i> (1)
<i>conocer</i>	<i>coñecer</i> (189)	<i>conecer</i> (1) e <i>coñocer</i> (1)
<i>ingeniero</i>	<i>enxeñeiro</i> (1)	<i>enxeniero</i> (1) e <i>inxenheiro</i> (1)
<i>exprimir</i>	<i>espremer</i> (0)	<i>expremir</i> (1)
<i>ingeniería</i>	<i>enxeñería</i> (0)	<i>inxeniería</i> (1)
<i>paladar</i>	<i>padal</i> (5)	<i>padar</i> (1)
<i>poseer</i>	<i>posuír</i> (33)	<i>posúer</i> (1)
<i>sumergir</i>	<i>somerxer</i> (2)	<i>sumerxer</i> (1)
<i>surgir</i>	<i>xurdir</i> (35)	<i>xurxir</i> (1)

Cómpre sinalar a presenza de distintas variantes para un mesmo concepto, como *(des)vantaxa* e *(des)ventaxe*, que se suma á xa mencionada *(des)ventaxa*; *inxenheiro*, *enxeniero* e a forma citada máis arriba *inxenieiro*, así como *conecer* e *coñocer*, as cales poden ser tamén variantes dialectais de orixe galega (*vid.* §3.2.1.5). Tal e como se pode ver na Táboa 41 do Apéndice, a maioría das formas teñen escasa ou nula presenza nos corpus (*desvantaxa*, *expremir*, *objecto* e *padar* non se recollen en ningún texto). As excepcións son *conecer*, *coñocer*, *costo* e *ventaxe*, con máis de 20 exemplos en cando menos un dos dous corpus analizados. No *TLPGP* non se rexistra ningunha forma agás *conecer*.

¹⁴ *Costo* podería ser tamén unha importación directa do español, equivante a *coste*, pero de acordo co *Diccionario panhispánico de dudas* da Real Academia Española (2005, s.v. *costo*) é unha forma máis utilizada no español de América.

Finalmente, incluímos tamén no grupo de voces adaptadas a voz *se*, recollida no fragmento «Se que é certo que grazas a importancia que acadou esta tamén a levar un cambio na sociedade». Neste caso, emprégase a forma *se* en vez de *si* probablemente pola extensión da equivalencia cast. *si* / gal. *se* que atopamos na conxunción condicional a estoutro contexto, onde o que debería figurar é a partícula afirmativa *si*.

De acordo co exposto e cos datos recollidos no Apéndice, entre as formas híbridas analizadas neste apartado hai tanto voces ou expresións ben testemuñadas no galego, sexa no galego popular, no galego elaborado ou en ambos, como é o caso de *parexa*, *sen embargo*, *mentras* ou *calquer*, que poderían ser préstamos integrados no repertorio mental das persoas autoras dos textos, como outras formas que non se rexistran ou apenas o fan nos textos galegos, como ocorre con *aconsexable*, *chear* ou *dineiro*, e que poden resultar de adaptacións *ad hoc* de palabras do castelán realizadas sobre todo por alumnado castelanfalante (interferencias léxicas de acordo coa terminoloxía presentada en §3.2.1.1).

3.2.1.3. Hipergaleguismos

Neste apartado presentamos un grupo de 110 formas (45 distintas) en que se realiza unha sobrexeneralización dun padrón de equivalencias galego-castelán (Dubert, 2004, p. 239) que dá lugar á modificación dunha palabra ou dunha secuencia que é coincidente en galego e castelán. Así, por exemplo, aplicando padróns do mesmo tipo que vimos no apartado precedente, *piscina*, forma estándar en galego e castelán, pasa a *pisciña*. O 3,66% dos códigos de orixe asignados ás palabras que presentan diferenzas coa voz estándar que van máis alá do xénero e da acentuación pertencen a este grupo.

As formas analizadas neste apartado adoitan caracterizarse como *hiperenxebrismos* ou *hipergaleguismos*, termos que con frecuencia se vinculan co diferencialismo característico do galego escrito da primeira metade do século xx. De acordo co que sinala Monteagudo (2003, p. 55) en relación coas persoas cultivadoras do galego do período de entreguerras que estaban conformando o galego culto,

Todo vocábulo formalmente idéntico ou similar en galego e castelán (*paronomasia* ou *cognato xémeo*), sexa culto ou patrimonial, é obviado ou é suxeito a un tratamento diferenciador. Un procedemento diferencialista moi habitual consiste na aplicación analóxica de regras de conversión fonolóxica do castelán ao galego, inducidas a partir da comparanza entre series de palabras de ambos os idiomas: a palabra *estraño* trastócase en *estrano*, segundo o paralelo cast. *año* / galego *ano*; a palabra *brillar* trócase en *brilar*, en paralelo a cast. *villa* / gal. *vila*; *oficina* convértese en *oficiña*, segundo paralelo do cast. *harina* / gal. *fariña*; *silencio* tórnase

silenzo, conforme o paralelo do cast. *precio*/ gal. *prezo*, etc. En definitiva, os cognatos xémeos son ultracorrrixidos e o resultado son formas *hiperenxebres*.

Evidentemente, as/os estudantes que escriben os textos de *CORTEGAL* procuran empregar formas estándar, evitando deste xeito ser sancionadas/os, de tal xeito que a presenza dos hipergaleguismos rexistrados no corpus non se pode explicar polo afán diferencialista con respecto ao castelán que atopabamos na primeira metade do século xx, onde aínda non existía unha norma establecida, que precisamente se estaba a construír: en *CORTEGAL*, tales voces son seleccionadas por se considerar que forman parte do repertorio normativo. Isto non é óbice para que os procesos comparativos entre galego e castelán que subxacen á creación das formas sexan os mesmos que atopabamos no século pasado e mesmo algúns dos exemplos proporcionados por Monteagudo na cita precedente coinciden cos atopados en *CORTEGAL* (*estrano* e *oficiña*).

Do mesmo xeito que fixemos no apartado anterior, clasificaremos as voces atopadas de acordo co padrón de equivalencias que explica o cambio, comezando por aquelas modificacións que afectan a consoantes, que se ofrecen na Táboa 17.

TÁBOA 17. Modificacións que afectan a elementos consonánticos

Padrón		Ocorrencias		Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma estándar	Forma non estándar		
<n> intervocál.	<ñ>	<i>gaña</i> (18), <i>medicina</i> (3), <i>piscina</i> (0)	<i>gaña</i> (7), <i>mediciña</i> (2), <i>pisciña</i> (1)...	14	7
<n> intervocál.	--	<i>personaxe</i> (8), <i>abandonar</i> (7), <i>almacenar</i> (1)...	<i>persoaxe</i> (7), <i>abandoar</i> (5), <i>almacear</i> (1)...	15	5
oclusiva + <l>	oclusiva + <r>	<i>complicado</i> (9), <i>planta</i> (2), <i>accesible</i> (6)...	<i>compricado</i> (2), <i>pranta</i> (2), <i>acesibre</i> (1)...	7	5
<ñ> intervocál.	<n>	<i>estraño</i> (6)	<i>estrano</i> (3)	3	1

O grupo máis numeroso é o conformado por palabras en que cae unha consoante, concretamente un <n> intervocálico, seguindo o esquema de *luna* / *lúa* ou *freno* / *freo*. Ocorre en *persoaxe*, onde tamén se pode apelar á influencia do substantivo *persoa* (vid. §3.2.1.4), *abandoar*, *almacear*, *destroar*¹⁵ e *zoa*.

¹⁵ A interpretación desta forma é dubidosa. A/o estudante escribiu *destrou* (a forma correcta de *destroar* sería *destrouu*). O contexto é o seguinte: «a chegada da pataca destrou a castaña».

Ao seu lado temos casos en que o <n> intervocálico pasa a <ñ>. De acordo con correspondencias como *camino* / *camiño* ou *vino* / *viño*, encontramos *mediciña*, *destiñar*, *oficiña*, *pisciña* e *vaciña*. Ademais, a equivalencia <n> / <ñ> esténdese máis alá de palabras rematadas en *-ino* / *-ina*, ou derivadas destas, posto que tamén rexistramos *gaña* e *mañeira*¹⁶. Neste último caso, cómpre sinalar a existencia dun certo número de correspondencias *-inero* / *-iñeiro* ou *-inera* / *-iñeira*, sobre todo derivados de formas en *-ino* / *-iño*, *-ina* / *-iña* (*marinera* / *mariñeira*, *cocinera* / *cociñeira*, *molinera* / *muiñeira*...), que poden estar na base da modificación realizada. No caso de *gana*, podería quizais tamén apelarse á influencia do verbo *gañar*.

Por outro lado, tamén atopamos un caso en que a modificación se produce no sentido contrario: seguindo o modelo de *año* / *ano*, *caña* / *cana*..., <ñ> pasa a <n>, dando lugar a *estrano*.

Finalmente, outra modificación consonántica ten lugar no paso de lateral a golpeada en grupo consonántico iniciado por oclusiva: rexistramos *compricado*, *pranta*, *acesibre*, *notablemente* e *frotar*, que seguen modelos como *plata* / *prata* ou *flecha* / *frecha*.

Noutros casos os elementos afectados son vocálicos, tal e como se pode ver na Táboa 18.

TÁBOA 18. Modificacións que afectan a ditongos

Padrón		Ocorrencias		Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma estándar	Forma non estándar		
<io>	<o>	<i>comercio</i> (26), <i>desprestixio</i> (0), <i>desperdicio</i> (4)...	<i>comerzo</i> (3), <i>desprestixo</i> (2), <i>desperdizo</i> (1)...	11	8
<ia>	<a>	<i>influencia</i> (33), <i>apreciar</i> (26), <i>convivencia</i> (3)...	<i>influenza</i> ¹⁷ (6), <i>aprezar</i> (2), <i>convivenza</i> (1)...	12	6
<ie>	<e>	<i>ingrediente</i> (28), <i>concienciar</i> (15), <i>inconveniente</i> (6)	<i>ingredente</i> (5), <i>concienciar</i> (1), <i>inconvenente</i> (1)	7	3

¹⁶ Non incluímos aquí os casos de *gañanza*, *gañancia* (gal. estándar e esp. *ganancia*) e *expoñencialmente* (gal. estándar e esp. *exponencialmente*), cuxa nasal palatal se explica moi probablemente por estaren creadas sobre *gañar* e *expoñente*. Vid. o apartado 3.2.1.4.

¹⁷ A forma *influenza* existe na proposta normativa do galego como sinónimo de *gripe*, pero é unha palabra pouco frecuente e semanticamente desvinculada do contido 'influencia', motivo polo cal etiquetamos as ocorrencias de *influenza* no nivel léxico e non no semántico.

Nos tres casos ofrecidos na Táboa 18 encontramos exemplos de monotongación de ditongos. Un grupo numeroso está conformado por palabras en que se modifica a secuencia <cia> ou <cio>, pasando a <za> e <zo> respectivamente, seguindo o modelo de *servicio* / *servizo* ou *gracia* / *graza*: na Táboa 42 do Apéndice poden verse todos os exemplos. No caso de *aprezar* e *desgrazadamente* é máis que probable tamén a influencia das formas da súa familia *prezo* / *desprezo* / *menosprezo* e *graza* / *desgraza* respectivamente, mentres que para *convivenza*, *influenza*, *referenza* e *vivenza* se pode apelar tamén a unha formación sobre os verbos co alomorfo-*nza* (*vid.* para unhas e outras §3.2.1.4). Cómpre mencionar, por outra banda, as formas *desprestixo*, *prestixo* e *prestixoso* (as dúas ocorrencias de *desprestixo* son proporcionadas pola mesma persoa, e os exemplos *prestixo* e *prestixoso* por outra diferente). Aínda que non existen correlacións entre as secuencias <gio>, <gia> / <xo>, <xa> en castelán e galego, é posible que a correlación <cio>, <cia> / <zo>, <za> arriba sinalada, apoiada por outros casos como *vendimia* / *vendima* ou *bestia* / *besta*, explique a modificación realizada, do mesmo tipo ca a que encontramos en *versonar*¹⁸.

En tres casos encontramos a monotongación do ditongo <ie>, que pasa a <e>, seguindo o modelo de *pedra* / *pedra*, ou *ardiente* / *ardente*: *ingrediente*, *concienciar* e *inconveniente*. As formas acabadas en *-ente* tamén se poderían incluír na Táboa 19, que se ofrece a seguir, en que recollemos aqueles casos en que as modificacións afectan a terminacións, pois podería apelarse ao padrón <-iente> / <-ente> que rexistramos, por exemplo, en *pendiente* / *pendente* ou *corriente* / *corrente* para explicar a súa formación. Tamén se poderían incluír nesta táboa as voces rematadas en *-nza* que acabamos de comentar.

TÁBOA 19. Modificacións que afectan a terminacións

Padrón		Ocorrencias		Total de formas non estándar	Total de formas non estándar distintas
Castelán	Galego	Forma estándar	Forma non estándar		
<-ero, a>	<-eiro, a>	<i>sincero</i> (0), <i>sinceramente</i> (9), <i>mero</i> (16)...	<i>sinceiro</i> (2), <i>sinceiramente</i> (2), <i>meiro</i> (1)...	31	5
<-iente>, <-yente>	<-inte>	<i>influyente</i> (0), <i>ingrediente</i> (28), <i>nutriente</i> (4)	<i>inflúinte</i> (5), <i>ingredinte</i> (1), <i>nutrinte</i> (1)	7	3
<-ano, -ana>	<-án, á>	<i>humano</i> (113), <i>semana</i> (14)	<i>humán</i> (2), <i>semá</i> (1)	3	2

¹⁸ Aínda que no *CORGA* e no *TILG* non hai exemplos de *versonar*, si os hai de *versón*.

Por un lado, cómpre citar uns poucos casos en que a terminación *-era*, *ero* pasa a *-eira*, *eiro*: *sinceiramente*, *sinceiro*, *austeiro* e *meiro*. Aos exemplos indicados pódese engadir *calqueira*, que xa foi presentado entre os calcos (§3.2.1.2), pero que tamén se pode explicar pola adaptación da terminación *-era* a *-eira*.

Por outro lado, tamén atopamos dous exemplos de evolución da terminación *-ano*, *-ana* a *-án*, *á*. Trátase de *humán* e *semá*, que seguen o esquema de *aldeano*, *-na* / *aldeán*, *-á* ou *artesano*, *-na* / *artesan*, *-sá*. Con todo, *semá* pode ser tamén unha variante dialectal localizada nalgúns puntos de Galicia de acordo cos datos do *Atlas lingüístico galego* (en diante *ALGa*) e do *TLPGP* (*vid.* §3.2.1.5).

Finalmente, cómpre mencionar as palabras *inflúinte*, *ingredinte* e *nutrinte*, que se poden explicar polas correlacións do tipo de *siguiente* / *seguinte* ou *saliente* / *saínte* que encontramos nalgunhas parellas na comparación castelán / galego. Con todo, no caso de *inflúinte* (o único dos hipergaleguismos recollido na Táboa 7, cunha presenza do 100%, posto que non hai exemplos de *influyente*) e de *nutrinte*, podemos tamén apelar a unha formación regular sobre os verbos *influir* e *nutrir* seguindo o modelo de formas galegas como as mencionadas, sen necesidade de apelar a unha comparación co español (*vid.* §3.2.1.4).

Aínda que se adoita falar de ultracorrección para casos como os que estivemos a ver (Monteagudo, 2003, p. 55; Kabatek, 2000, p. 33) e, malia que, en realidade, esta interpretación de hipercorrección dunha voz *galega* é sempre sostible (e o código empregado na nosa etiquetaxe, *L_hc*, correspóndese con ela), na maior parte dos casos tamén se podería falar simplemente dun mecanismo de adaptación desde o castelán do mesmo tipo do que vimos en §3.2.1.2. Nótese que a maior parte das palabras modificadas son coincidentes en galego e castelán (*piscina*, *ingrediente*, *zona*, *comercio*...). Nestes casos, en realidade, desde a perspectiva da persoa falante, podemos supoñer que operan os mesmos mecanismos de adaptación sinalados en §3.2.1.2. Pártese dunha palabra identificada no repertorio lingüístico como castelá, pero non como galega estándar, porque se non, non se modificaría (por exemplo, *empleo* en §3.2.3.2 e *complicado* no apartado actual), e, aplicando un padrón de equivalencias castelán / galego, lévase a cabo unha adaptación desa forma na procura da voz galega estándar.

Existen, con todo, uns poucos casos en que a coincidencia plena entre castelán e galego non existe: *calqueira* (esp. *cualquiera* / gal. *calquera*), *desprestixo* (esp. *desprestigio* / gal. *desprestixio*), *estrano* (esp. *extraño* / gal.

estraño), *inflúinte* (esp. *influyente* / gal. *influyente*) *mañeira* (esp. *manera* / gal. *maneira*), *persoaxe* (esp. *personaje* / gal. *personaxe*), *prestixo* (esp. *prestigio* / gal. *prestixio*), *prestixoso* (esp. *prestigioso* / gal. *prestixioso*), *razonamento* (esp. *racionamiento* / gal. *racionamento*) e *vaciña* (esp. *vacuna* / gal. *vacina*). Na maior parte dos casos, podería tamén considerarse o castelán como punto de partida, con adaptación dobre (desde *prestigio* confórmasse *prestixo* pasando a velar a postalveolar e reducindo <io> a <o>), pero en *calqueira*, *vaciña* e talvez *estrano*¹⁹ non parece posible supoñer unha adaptación baseada en padróns de correlación a partir das formas do español *cualquiera*, *vacuna* (e *extraño*). Nestes casos, só cabe postular unha ultracorreción desde as formas do galego *calquera*²⁰, *vacina* (e *estraño*), nas que se identifican secuencias (terminacións <-era>, <-ina> e <ñ> intervocálico) que se consideran sospeitosas de non ser galegas, levando a cabo a súa modificación.

Debe terse en conta, finalmente, que a carón dalgunhas formas con nula ou escasa presenza no textos escritos, tal e como se recolle na Táboa 42 do Apéndice, como *concenciar*, *desperdizo*, *ingredinte* ou *meiro*, e que resultan probablemente de adaptacións *ad hoc* feitas polo alumnado á hora de redactar os textos, algúns hipergaleguismos que xurdiron no período de entreguerras ou mesmo antes tiveron certo éxito na lingua formal posterior (*abandoar*, *calqueira*, *estrano*, *gaña*, *humán*, *persoaxe*, *pranta*...), de tal xeito que poderían formar parte do repertorio léxico das/dos estudantes que os empregaron (véxase na Táboa 42 do Apéndice a presenza desas e outras formas no *CORGA*, que só recolle textos de 1975 en diante). No que respecta á presenza no *TLPGP*, as únicas que se rexistran son *calqueira*, *mediciña* e *oficiña*²¹ (nun concello en cada caso) e *semá* (en dous).

3.2.1.4. Formas creadas por analoxía a partir doutras voces galegas

Integran este grupo 98 palabras correspondentes a 55 formas diferentes etiquetadas en *CORTEGAL* con L_anal (supón un 3,26% dos códigos de orixe asignados a palabras con diferenzas con respecto á forma estándar que van máis alá do xénero e da acentuación). As formas deste grupo, que clasificaremos tendo en conta características do seu proceso formativo, poden

¹⁹ Dado que hai varios casos en que á secuencia inicial <ex-> do español lle corresponde <es-> en galego (*extranjero* / *estranxeiro*, *extender* / *estender*...), tamén se podería falar aquí dunha dobre adaptación desde *extraño*, co paso de <ex-> a <es-> e de <ñ> a <n>.

²⁰ Lembrese, en calquera caso, que *calqueira* tamén se pode explicar como un calco estrutural de *cualquiera*, tal e como comentamos en §3.2.1.2.

²¹ A palabra *oficiña* rexístrase na mesma obra, sobre a fala de Cangas do Morrazo, en que se recolle *calqueira*. Indicase que aparece esporadicamente (González González, 2003, p. 557).

conibirse como o resultado dunha creación dentro do propio galego a partir de analoxías con outras formas desta lingua (aínda que tamén poida estar presente por tras a influencia dunha voz do castelán). Dúas delas, *influínte* e *inconvinte*, aparecen recollidas na Táboa 7 por presentaren unha porcentaxe de aparición superior ao 50% no total do concepto.

Por un lado, agrupamos na Táboa 20 un conxunto de 33 ocorrencias, correspondentes a 11 voces distintas, cuxa forma se pode explicar por un proceso de analoxía con outros vocábulos da familia léxica ou de familias formalmente próximas.

TÁBOA 20. Formas creadas por analoxía con outras voces da familia de palabras ou de familias formalmente próximas

Forma non estándar	Forma estándar
<i>inconvinte</i> (8)	<i>inconveniente</i> (6)
<i>gaña</i> (7)	<i>gana</i> (18)
<i>persoaxe</i> (6)	<i>personaxe</i> (8)
<i>xeralizar</i> (3)	<i>xeneralizar</i> (7)
<i>aprezar</i> (2)	<i>apreciar</i> (26)
<i>gañanza</i> (2)	<i>ganancia</i> (11)
<i>desgrazadamente</i> (1)	<i>desgraciadamente</i> (3)
<i>expoñencialmente</i> (1)	<i>exponencialmente</i> (3)
<i>gañancia</i> (1)	<i>ganancia</i> (11)
<i>inconvente</i> (1)	<i>inconveniente</i> (6)
<i>obsolescencia</i> (1)	<i>obsolescencia</i> (12)

Deixando a un lado o caso de *gana*, que xa foi comentado en §3.2.1.3, e que podería explicarse por influencia do verbo *gañar*²², aínda que ambas as voces pertencen a familias de palabras distintas, no resto dos casos encontrámonos con vocábulos cuxa forma se pode explicar a partir dun proceso de analoxía con voces patrimoniais ou máis sinxelas da correspondente familia léxica: dado que o verbo é *poñer* (e a partir del se forma a cadea *expoñer*, *expoñente*), a/o estudante escribe *expoñencialmente*; a partir de *xeral*, constrúese

²² Tanto o verbo *gañar* como o verbo *ganar*, de idéntico significado, son estándar en galego. Con todo, na lingua escrita é máis frecuente o verbo *gañar*. Se analizamos os datos do TILG, no período 2000-2013 (último ano do que hai datos), encontramos 582 exemplos de *gañar* fronte a 78 de *ganar*. No CORGA, desde o ano 2000 ata 2020, o derradeiro que se ofrece, os resultados son aínda máis dispares: 5513 casos de *gañar* fronte a 156 de *ganar*. Con respecto a CORTEGAL atopamos 197 exemplos de *gañar* fronte a 11 de *ganar*. De feito, a partir da nosa experiencia docente, podemos afirmar que parte do alumnado considera que *ganar* non é unha forma galega estándar.

a forma perfectamente regular desde o punto de vista formal e semántico, *xeralizar*; a ausencia do <n> en *personaxe* pode explicarse por influencia do substantivo *persoa*; as formas *gañancia* e *gañanza* responden á creación a partir da variante con nasal palatal *gañar* e dúas variantes posibles para o afixo (-anza, como en *confianza*, ou -ancia, como en *tolerancia*); *inconvinte* e *inconvente* son o resultado da formación sobre *convir* de acordo con dúas variantes alomórficas existentes no galego para as formas creadas sobre os verbos da terceira conxugación (en *inconvinte*, seguindo o modelo de *seguinte* ou *ferinte*, a vogal temática é <i>, mentres que en *inconvente*, seguindo o modelo de *servente* ou *vivente*, a vogal seleccionada é <e>); *desgrazadamente* responde á analogía con *graza* e *desgraza* e *aprezar* resulta coherente con *prezo* e *prezar*. Finalmente, *obsolescencia* é unha forma non rexistrada nas fontes presentadas no Apéndice (Táboa 43), que podería responder simplemente a un lapsus. Con todo, cómpre sinalar que se encontran varios exemplos na rede (por exemplo, «coñecerán os síntomas que padece o planeta, como lle afectan os cambios, distintos tipos de contaminación, a *obsolescencia* planificada e medidas que podemos adoptar para o coidado do noso planeta»²³) e que a presenza das formas rexistradas en *CORTEGAL* e na rede talvez se poida vincular co adxectivo *obsoleto* e coa interpretación de *obso-* como raíz, de xeito que *obsolescencia* viría a regularizar a familia léxica.

Debe repararse tamén no feito de que algunhas das voces da Táboa 20 (*gaña*, *persoaxe*, *aprezar*, *desgrazadamente*) xa foron mencionadas no apartado precedente, como posible resultado da aplicación de padróns de equivalencia galego-castelán. Cómpre destacar, por último, a convivencia nalgúns casos de varias solucións para unha mesma forma, como consecuencia da coexistencia de diferentes posibilidades morfolóxicas no galego: *inconvinte* e *inconvente* por un lado, *gañancia* e *gañanza* por outro.

En segundo lugar, na Táboa 21, incluímos varios exemplos en que se percibe influencia de elementos do paradigma flexivo. Rexistramos aqueles casos en que se produce un cambio na raíz verbal e/ou na conxugación dun verbo e tamén o pronome neutro *elo*, cuxa forma pode explicarse por analogía coas formas femininas do seu paradigma *ela* e *elas*.

²³<https://www.comarcasnarede.com/comarcas/val-minor/gondomar/sos-a-terra-nova-actividade-medioambiental-en-gondomar/> [última consulta: 04/03/2023].

TÁBOA 21. Formas creadas por analogía con outras voces do paradigma flexivo

Forma non estándar	Forma estándar
<i>elo</i> (10)	<i>iso, isto</i> (--)
<i>entreteñemento</i> (2)	<i>entretemento</i> (17)
<i>teñer</i> (2)	<i>ter</i> (401)
<i>conseguer</i> (1)	<i>conseguir</i> (78)
<i>consomer</i> (1)	<i>consumir</i> (330)
<i>descobrir</i> (1)	<i>descubrir</i> (15)
<i>locer</i> (1)	<i>lucir</i> (1)
<i>manteñer</i> (1)	<i>manter</i> (57)
<i>seguer</i> (1)	<i>seguir</i> (86)
<i>sosteñer</i> (1)	<i>soster</i> (4)

Antes de comentar os datos da táboa, cómpre puntualizar dúas cuestións. En primeiro lugar, con respecto á mudanza de conxugación, unicamente se consideran dentro do nivel léxico (e se presentan na táboa) os infinitivos, posto que neles se pode deducir claramente o cambio de conxugación verbal. Así, por exemplo, é considerada desviación léxica a forma *seguer*, onde se emprega a vogal temática propia da segunda conxugación, pero non a forma *sega*, que se pode explicar simplemente por unha regularización da vogal radical no presente de subxuntivo do verbo *seguir*, e que consideramos no nivel morfolóxico. Tamén se analizan neste último nivel formas como *seguera* (por *seguera*) ou *consequemos*, posto que, aínda que nestes casos se podería interpretar que se produce un cambio de conxugación, dado que figura nelas a vogal temática <e> e non <i>, propia nesos tempos e persoas da segunda conxugación, é tamén posible apelar a unha selección inadecuada desta última nesas formas concretas, sen que isto implique mudanza de paradigma conxugacional. De feito, a persoa que escribe *consequemos* utiliza no mesmo texto *conseguir*. Nótese que a conxugación destes verbos no galego estándar, con <e> na P2, P3 e P6 do presente de indicativo (*segues, segue, seguen*) en vez do <i> que atopamos no castelán e en boa parte do galego popular, adoita ofrecer certas dificultades ao estudantado (tal e como se pode comprobar no propio *CORTEGAL*) e as dúbidas xeradas poden dar lugar á extensión da diferenza <e> / <i> a contextos que non lles corresponden, mesmo máis alá da vogal radical.

Con respecto á mudanza na raíz, unicamente consideramos como desviacións léxicas aquelas que afectan tamén a infinitivos (ou derivados deverbais), por ser o tema que encontramos neles o punto de partida de novas

creacións léxicas (por exemplo, se temos *descobrir*, o derivado será *descobremento*). Outras mudanzas, como a que rexistramos en formas finitas como *consoman*, son consideradas extensións das alternancias vocálicas propias destes verbos, sen que necesariamente teñan repercusións a nivel léxico (*consoman* pode ser unha forma do verbo *consumir*, en que a alternancia vocálica que atopamos na P2, P3 e P6 do presente de indicativo, *consomes*, *consume*, *consomen*, se estende ao subxuntivo), de tal xeito que son consideradas desviacións da dimensión morfolóxica.

Indo xa aos datos da táboa, debemos lembrar que *elo* foi tamén presentado en §3.2.1.2 como unha posible adaptación desde o castelán, con paso de <ll> a <l>; en calquera caso, responda a un proceso de adaptación, de anaxoxía ou a ambos, está clara a influencia da forma do español na súa creación.

Por outro lado, en *entreteneñemento* (sobre *entreteneñer*), *teñer*, *manteñer* e *sosteñer*, mencionadas tamén en §3.2.1.2 como resultado dunha posible adaptación ao galego desde o castelán, pode postularse ademais a influencia doutras formas do paradigma, como a P1 (*teño*) e a P6 (*teñen*) do presente de indicativo e todo o presente de subxuntivo (*teña*, *teñas*...) para explicar, nun proceso de regularización, a presenza da nasal palatal.

No caso de *consomer*, *descobrir* e *locer*, de súa vez, encontrámonos ante un cambio de vogal radical moi posiblemente por influencia das formas regulares da P2, P3 e P6 do presente de indicativo *consomes*, *consume*, *consomen*; *descobres*, *descobre*, *descobren* e *loces*, *loce*, *locen* respectivamente. Esta mudanza de vogal radical vai acompañada dun cambio na conxugación no caso de *consomer* e *locer*, cambio que tamén ten lugar en *seguer* e *conseguer*, cuxa forma probablemente se poida explicar igualmente a partir das formas da P2, P3 e P6 do presente de indicativo (*con*)segue, (*con*)segues, (*con*)seguen: a maior parte dos verbos da terceira conxugación que teñen <e> como vogal radical no infinitivo presentan un <i> na P1, P2, P3 e P6 do presente de indicativo (por exemplo, no caso de *medir*, *mido*, *mides*, *mide*, *miden*). Só un grupo reducido de verbos da terceira conxugación, entre os que se atopan *seguir* e os seus derivados, teñen <e> na vogal radical da P2, P3 e P6, de tal xeito que a presenza desta vogal en tales formas é moito máis habitual en verbos da segunda conxugación, o que pode conducir á formación de (*con*)seguer. Razóns de frecuencia tamén poden explicar o cambio de conxugación en *consomer* e *locer*: a vogal radical <o> é moito máis habitual en verbos da segunda conxugación (moitos deles, ademais, moi frecuentes, como *poñer*, *correr*, *morrer*...) ca en verbos da terceira, o que pode explicar que a partir de *loce* ou *consume* se constrúan *locer* e *consomer*.

Un terceiro grupo de formas, recollidas na Táboa 22, está constituído por aquelas palabras en que a única diferenza coa forma estándar radica na selección dun alomorfo non estándar dun sufixo.

TÁBOA 22. Formas creadas con alomorfos dos sufixos estándar

Forma non estándar	Forma estándar
<i>influenza</i> (6)	<i>influencia</i> (33)
<i>influínte</i> (5)	<i>influyente</i> (0)
<i>convivenza</i> (1)	<i>convivencia</i> (3)
<i>nutrinte</i> (1)	<i>nutriente</i> (4)
<i>referenza</i> (1)	<i>referencia</i> (14)
<i>vivenza</i> (1)	<i>vivencia</i> (2)

Por unha banda, atopamos varios casos de formas que presentan o alomorfo *-nza* en vez de *-ncia* (*convivenza*, *influenza*, *referenza* e *vivenza*) e que poderían ser o resultado de creacións regulares sobre os verbos *convivir*, *influír*, *referir* e *vivir* respectivamente, de acordo co modelo doutras formas que teñen o alomorfo *-nza* como *crenza*, *herdanza*, *doenza* etc. Nótese que estas formas xa foron mencionadas no apartado 3.2.1.3, como posible resultado tamén dunha hiperoxeneralización do padrón <cia> / <za>.

En segundo lugar, encontramos dous casos en que se produce unha selección non estándar da vogal temática que acompaña o sufixo <-nte>. Trátase de *influínte* e *nutrinte*, que poden ser o resultado da formación regular sobre *influír* e *nutrir* seguindo o modelo de *seguinte* ou *ferinte*, onde a vogal temática é <i> (Dubert, 2004, p. 241). Nótese que as formas *influínte* e *nutrinte* tamén se mencionaron como posibles casos de hipergaleguismos (*vid.* §3.2.1.3).

O seguinte subgrupo de formas supón a creación de palabras seguindo pautas morfolóxicas do galego e en todas elas encontramos procesos de afixación. Trátase de palabras posibles nesta lingua, pero non rexistradas nos dicionarios, porque contan cunha «competidora» coa mesma raíz, pero diferente(s) afixo(s) (ou ausencia de afixos). Todas elas aparecen rexistradas na Táboa 23.

TÁBOA 23. Formas creadas mediante sufixación con competidores coa mesma raíz

Forma non estándar	Forma estándar
<i>asombrante</i> (1)	<i>asombroso</i> (1)
<i>bolsátil</i> (1)	<i>bolsista</i> (0)
<i>diferenciativo</i> (1)	<i>diferencial</i> (0)
<i>divulgamento</i> (1)	<i>divulgación</i> (5)
<i>emplataxe</i> (1)	<i>empratado</i> (2)
<i>estourido</i> (1)	<i>estouro</i> (0)
<i>famosidade</i> (1)	<i>fama</i> (64)
<i>inmemorable</i> (1)	<i>inmemorial</i> (0)
<i>irrecoñecido</i> (1)	<i>descoñecido</i> (17)
<i>lalinés</i> (1)	<i>lalinense</i> (0)
<i>monotonidade</i> (1)	<i>monotonía</i> (2)
<i>sacrificación</i> (1)	<i>sacrificio</i> (1)
<i>sacrificante</i> ²⁴ (1)	<i>sacrificado</i> (0)
<i>vanguardoso</i> (1)	<i>vanguardista</i> (4)

Como se pode observar, aínda que poida haber nalgúns casos algúns matices de significado diferentes en función do afixo, as palabras das dúas columnas son semanticamente (quasi)equivalentes entre si e todas elas están ben conformadas de acordo coas regras morfolóxicas que operan no galego.

No caso de *inmemorable*, debe terse en conta que, aínda que escasas, existen unhas poucas formas construídas co sufixo *-ble* sobre bases nominais, como *favorable* ou *saudable*. En calquera caso, *inmemorable* resulta probablemente dun intento de motivación formal da voz *inmemorial*. O significado desta palabra, ‘que é tan antigo que non queda lembranza da súa orixe’ (definición extraída do dicionario *Digalego* s.v.), implica *imposibilidade* de recordar, de tal xeito que a presenza do sufixo *-ble*, que expresa posibilidade pasiva (Álvarez e Xove, 2002, p. 708), combinado co prefixo negativo, dálle todo o sentido á voz. *Estourido* é unha formación regular sobre o verbo *estourar* ‘estoupar’, creada seguindo o modelo de *estoupido* ou *estalido*, cun sufixo *-ido* habitual para conformar palabras que designan voces ou ruídos (cf. ademais das formas citadas, *ronquido*, *zunido*, *xemido*...). *Lalinés*, de súa

²⁴ Esta última forma aparece empregada nos corpus como termo de lingua de especialidade (da antropoloxía ou da psicopedagogía) e cun valor diferente ao de *CORTEGAL*, onde o significado é ‘duro, que implica sacrificio’. Por este carácter especializado, etiquetamos *sacrificante* no nivel léxico (como resultado dunha creación da/do estudante) e non no nivel semántico, como un emprego semanticamente inadecuado dunha palabra que posúe outros significados.

vez, tamén responde a un procedemento habitual na formación de xentilicios (cf. *vigués, francés, cangués...*). No que respecta á forma *bolsátil*, é moi probable a influencia da forma do español *bursátil*, do mesmo significado, de tal xeito que se combina a raíz regular *bols-* e o afixo seleccionado no castelán.

Noutro grupo menos numeroso, a forma creada tamén responde a pautas morfolóxicas do galego, pero desta volta a palabra non conta cun competidor da mesma raíz. Son os exemplos rexistrados na Táboa 24.

TÁBOA 24. Formas creadas mediante derivación sen competidores coa mesma raíz

Forma non estándar	Forma estándar
<i>acadamento</i> (1)	<i>consecución</i> (1)
<i>malgaste</i> (1)	<i>dispendio</i> (o)
<i>mobilístico</i> (1)	<i>dos móbiles</i> (1)
<i>recolleitar</i> (1)	<i>recoller</i> (7)

Mediante sufixación (*acadamento, mobilístico*) ou derivación sen afixos (Álvarez e Xove, 2002, p. 680) (*malgaste, colleitar*), créase un derivado sobre unha base verbal (*acadar, malgastar*) ou nominal (*móbil, colleita*), cubrindo deste xeito unha «lagoa» na correspondente familia de palabras. Con respecto a *mobilístico*, onde *móbil* se combina con *-ístico*, é probable a influencia de *automobilístico*, mentres que no caso de *recolleitar*, palabra que tamén se considera no grupo das transferencias desde o galego popular, posto que aparece testemuñada neste, é posible ademais a influencia do español *recolectar* á hora de explicar a conformación da voz.

A Táboa 25 está conformada por cinco palabras creadas pola / polo estudante mediante padróns do galego, pero cuxa forma non se corresponde co significado que aparentemente se quere expresar.

TÁBOA 25. Palabras cuxa forma non se corresponde co significado buscado

Forma non estándar	Forma estándar
<i>autoconsumismo</i> (1)	<i>consumismo</i> (518)
<i>contrapiar</i> (1)	<i>contrapeso</i> (o)
<i>innaturalidade</i> (1)	<i>antinaturalidade</i> (o)
<i>intercomercial</i> (1)	<i>de comercio internacional</i> (o)
<i>produtismo</i> (1)	<i>produtivismo</i> (2)

A seguir ofrécense os contextos de uso:

- (1) Por exemplo o consumo dos smartphones é un tema moi tratado hoxe en día isto aparece ligado ao *autoconsumismo* que levan a cabo algunhas persoas pola necesidade de adquirir estes aparellos
- (2) Este modelo económico basease no consumo individual. Pero a produción e o *contrapiar* deste
- (3) unha extrema exquisitez respecto á comida por parte de certos individuos, que chega a rozar extremos de *inaturabilidade*
- (4) Na actualidade vivimos nun mundo *intercomercial*, isto é, non resulta difícil atopar alimentos exóticos
- (5) A nosa vida social tamén está controlada polo *produtismo* e o consumismo

En *autoconsumismo*, de acordo co contexto, non se percibe a pertinencia semántica do elemento *auto-*, mentres que a análise do fragmento en que figura *innaturabilidade* desbota a presenza do contido de posibilidade transmitido polo sufixo *-bil-*. De súa vez, en *intercomercial* o prefixo *inter-* aplícase a un elemento inapropiado e falta a base con que se debería combinar (*nacional*). Finalmente, en *contrapiar* e *produtismo* o problema parece radicarse nas bases de derivación, que non se adecúan ao contexto de uso. No caso de *produtismo*, é posible que se trate dun lapso de escritura, posto que no mesmo texto figura a voz *produtivismo*.

Por outra banda, en dous exemplos, atopamos combinacións «imposibles» de bases e afixos. Son os casos rexistrados na Táboa 26.

TÁBOA 26. Combinacións «imposibles» de bases e afixos

Forma non estándar	Forma estándar
<i>moderidade</i> (1)	<i>moderación</i> (2)
<i>producidade</i> (1)	<i>produtividade</i> (198)

Nos dous casos parece combinarse un verbo (*moderar*, *producir*) co sufixo *-idade*, un sufixo que, no entanto, toma como base adxectivos. No primeiro substantivo, a forma estándar correspondente ten a mesma base e cambia o afixo (*-ción*), mentres que no segundo mantense o afixo pero varía a base (*produtivo*).

Finalmente, na Táboa 27, rexistramos outros tres casos cuxa forma talvez se podería explicar pola frecuencia de determinadas terminacións no galego:

TÁBOA 27. Voces cuxa forma se pode explicar pola recorrencia de certas terminacións

Forma non estándar	Forma estándar
<i>populariedade</i> (2)	<i>popularidade</i> (308)
<i>lonxetividade</i> (1)	<i>lonxevidade</i> (1)
<i>tercialización</i> (1)	<i>terciarización</i> (3)

Na voz *populariedade*, repetida dúas veces nun mesmo texto, figura o alomorfo *-edade* (en vez de *-idade*, que sería o que lle correspondería), precedido de <i>. É probable que esta terminación responda á analoxía co grupo de formas rematadas en <-iedade> (particularmente <-riedade>), onde o <i> forma parte da base de derivación (*notoriedade*, *precariedade*, *solidariedade*, *variedade*...). Poderíamos tamén apelar á extensión a novas formas da equivalencia galego-castelán *-(r)iedade* / *-(r)idad*, que encontramos en *solidariedade* / *solidaridad*, pero o feito de que esta correspondencia só afecte a unha parella de palabras e de que a forma *populariedade* se rexistre tamén en portugués, onde a ultracorrección con base no castelán non parece operar (por exemplo, «em esse período o índice de populariedade de VP aumentou», *Corpus do Português: Web/ Dialects*, Davies, 2016), levounos a descartar a opción de considerar tamén a palabra como un hipergaleguismo.

Lonxetividade é unha palabra rexistrada en *CORTEGAL*, da que non temos exemplos nos corpus de referencia, pero si na rede²⁵. Probablemente se produza por analoxía coas formas acabadas en <-tividade> (formadas polos sufixos *-tivo* e *-idade*, como *creatividade*, ou cos sufixos *-ivo* e *-idade* e unha base que contén <t>, como *deportividade*). Nótese que a maior parte das veces a secuencia <-vidade> está incluída na terminación <-tividade> (botando man dos lemas do *DRAG*, isto sucede aproximadamente en dous terzos das palabras que rematan en <-vidade>).

Finalmente, no caso de *tercialización*, voz da que rexistramos un exemplo no *CORGA* (vid. a Táboa 43 do Apéndice) e varios na rede, ademais dun lapso, podería supoñerse a influencia das numerosas formas rematadas en <-alización> (*-al+izar+ción*) (80 rexistradas no *DRAG*), moito máis numerosas ca as acabadas en <-arización> (18 no *DRAG*), ademais da posible perda para a/o estudante de vinculación semántica coa voz *terciario*.

Como se pode comprobar na Táboa 43 do Apéndice, agás *recolleitar*, ningunha das formas analizadas neste grupo se rexistran no *TLPGP*. No que respecta ao galego elaborado, algunhas teñen cando menos certa presenza

²⁵ Por exemplo, «xa naqueles anos falaba da súa lonxetividade polo tanto debe ter hoxe moitos, moitísimos anos» <https://www.rios-galegos.com/arb10.htm> [última consulta: 04/03/2023].

nos corpus: *aprezar, convivencia, descubrir*²⁶, *desgrazadamente, elo, estourido, gaña, gañancia, inconvinete, influenza, influínte, persoaxe, recolleitar, referenza, seguer, sosteñer e xeralizar* teñen máis de 20 ocorrencias en polo menos un dos dous corpus estudados e algunhas delas (*elo, estourido, influínte...*) son especialmente frecuentes. Un bo número das formas presentadas, en cambio, non están rexistradas en ningunha das fontes analizadas, entre as que se inclúen, como cabía esperar, todas aquelas cun proceso formativo irregular ou con inadecuación entre forma e significado, así como varias das que presentan un competidor coa mesma raíz (*famosidade, divulgamento...*), pero tamén outras, como é o caso de *inconvente* ou *consomer*. Así pois, unha vez máis, algunhas das formas rexistradas poden estar lexicalizadas no repertorio dalgún das/dos estudantes que as inclúen nos textos, mentres que outras probablemente resulten de creacións *ad hoc*.

3.2.1.5. *Transferencias do galego popular ou do galego elaborado*

Un total de 81 formas (30 distintas) foron consideradas neste grupo, en que se clasifican as palabras que, sendo formas propias do galego popular ou voces características do galego elaborado, por diferentes motivos (reducida extensión dialectal, voces consideradas arcaicas...) non están admitidas no código normativo. Un 2,70% dos códigos de orixe asignados ás palabras con desviacións que van máis alá do xénero e da acentuación corresponden a esta clase, etiquetada con L_gal. Deixamos a un lado as formas transferidas, con adaptación ou sen ela, do español e empregadas en calquera das variedades indicadas, así como os hipergaleguismos ou as formas creadas por analoxía con outras voces galegas que podemos atopar nalgúns textos do galego elaborado.

Por un lado, recollemos na Táboa 28 as voces galegas propias do galego popular, con indicación, ademais das formas equivalentes estándar e do número de ocorrencias en *CORTEGAL* dunhas e outras, da comisión delegada (CD) a que pertencen os textos en que se rexistran, así como da existencia ou non de correspondencia entre a distribución territorial da forma e cando menos algún dos concellos que abrangue a CD (a este respecto, *vid.* Álvarez de la Granja e Rodríguez Rodríguez neste volume, §2.1). Para a distribución territorial temos en conta en primeiro lugar os datos das diferentes enquisas do *ALGa* e, se neles non hai información ao respecto ou se interesa complementar, acudimos a outras fontes. Todas elas aparecen indicadas na última

²⁶ A elevada presenza de *descubrir* nos textos de tales corpus pode estar relacionada tamén co feito de ser esta a forma estándar do portugués. Por suposto, cabería tamén apelar a esta orixe no texto de *CORTEGAL* se a/o estudante que a proporcionou coñece esta lingua.

columna da táboa, onde GVGH representa o *Glosario de voces galegas de hoxe* de Constantino García, consultado a través do *Diccionario de dicionarios* (Santamarina, 2006).

TÁBOA 28. Voces non estándar de orixe galega presentes no galego popular

Forma non estándar	Forma estándar	CD	Coincidencia distrib. dialectal /CD	Fonte distrib. dialectal
<i>fruto</i> (4)	<i>froito</i> (8)	08, 16, 20, 23	Non	Mapas 330 e 331, vol. III, <i>ALGa</i>
<i>atrais</i> (3)	<i>atrás</i> (51)	14	Si	Ferreiro (1999, p. 354)
<i>bon</i> (3)	<i>bo</i> (409)	08, 12	Si	Mapa 256, vol. III, <i>ALGa</i>
<i>millor</i> (4)	<i>mellor</i> (358)	05, 19	Si	Mapa 138, vol. II, <i>ALGa</i>
<i>tar</i> (3)	<i>estar</i> (1609)	02, 06-07	Si	Mapa 256, vol. III, <i>ALGa</i>
<i>terneira/terneiro</i> (3)	<i>tenreira/tenreiro</i> (0)	07, 09, 11	Non	Mapas inéditos do <i>ALGa</i>
<i>fruta</i> (2)	<i>froita</i> (3)	03,05	Non	Mapas 330 e 331, vol. III, <i>ALGa</i>
<i>misturar</i> (2)	<i>mesturar</i> (3)	05,07	Si	<i>TLPGP</i> e GVGH
<i>pimento</i> (2)	<i>pemento</i> (2)	21	Non	Mapa inédito do <i>ALGa</i>
<i>buche</i> (1)	<i>boche</i> (0)	23	Si	<i>TLPGP</i>
<i>caír</i> (1)	<i>caer</i> (18)	04	Si	Mapa 160, vol. I, <i>ALGa</i>
<i>conecer</i> (1)	<i>coñecer</i> (189)	05	Non	Mapa inédito do <i>ALGa</i>
<i>coñocer</i> (1)	<i>coñecer</i> (189)	12	Non	Mapa 339, vol. III <i>ALGa</i>
<i>cuadro</i> (1)	<i>cadro</i> (2)	24	Non	Mapas 395 e 396, vol. III, <i>ALGa</i>
<i>cual</i> ²⁷ (1)	<i>cal</i> (242)	19	Si	Mapa 369, vol. II, <i>ALGa</i>
<i>cuanto</i> (1)	<i>canto</i> (153)	07	Non	Mapas 360, vol. II, <i>ALGa</i>
<i>tu</i> (1)	<i>ti</i> (34)	19	Si	Mapa 152, vol. II, <i>ALGa</i>
<i>millorar</i> (1)	<i>mellorar</i> (78)	05	Si	GVGH
<i>podrir</i> (1)	<i>podrecer</i> (1)	02	Si	Mapa 95, vol. I, <i>ALGa</i>
<i>recolleitar</i> (1)	<i>recoller</i> (7)	24	Non	<i>TLPGP</i> e GVGH
<i>semá</i> (1)	<i>semana</i> (14)	01-03-05	Non	Mapa 137, vol. IV, <i>ALGa</i> e <i>TLPGP</i>

²⁷ Non se inclúe neste grupo un segundo caso de *cual*, que vai precedido de *lo* (*lo cual*), por entenderse que neste caso se trata claramente dunha transferencia desde o español.

Como se pode observar, en pouco máis da metade dos casos existe correspondencia entre a extensión que ocupa a comisión delegada e a distribución xeográfica recollida nas fontes. Non ocorre así con formas que teñen asignada ademais outra orixe diferente, como é o caso de *cuanto* e *cuadro*, que poden ser o resultado tanto da conservación do grupo latino inicial [kw] seguido de [a] acentuado, propia dunha área do sur e leste de Galicia, como unha transferencia directa do español, *terneira/terreiro* e *pimento*, considerados tamén como posibles transferencias adaptadas e *conecer* e *coñocer*, presentados nas transferencias con mestura entre castelán e galego. Dada a inexistencia de correspondencia xeográfica, é bastante probable que a orixe destas formas sexa o español. Tampouco existe correspondencia no caso de *recolleitar*, voz rexistrada nalgúns concellos no TLPGP e no GVGH e que tamén se explicou como o resultado dunha analoxía, así como en *semá*, rexistrada nun punto no ALGa e en dous no TLPGP, interpretada así mesmo como un posible hipergaleguismo. En calquera caso, e no sentido contrario, a existencia de correspondencia tampouco descarta en absoluto a outra posibilidade explicativa. Así, por exemplo, a pesar de que a forma *cual* se rexistra nunha CD con concellos en que se mantén o elemento vocálico velar (sur de Pontevedra), isto non descarta que no texto concreto en que aparece a forma esta teña a súa orixe nunha transferencia do español dun castelanfalante habitual.

No que respecta a *fruta* / *fruto*, cómpre sinalar que non temos en conta a distribución recollida no mapa 6 do vol. III do ALGa para *fruta*, onde se ve que esta forma, moi probablemente por influencia do español, se emprega na maior parte de Galicia, e que *froita* e *fruita* se rexistran en moi poucos puntos, frecuentemente convivindo con *fruta*. Cinguímonos á zona galega en que o grupo latino -ŮCT- deu lugar a <u> (*luto*, *truta*), que ocupa unha pequena área no bloque oriental (Fernández Rei, 1990, pp. 48-51). Se atendemos a esta zona, non hai correspondencia entre as CD e a distribución territorial desta solución.

Como se pode verificar na Táboa 44 do Apéndice, todas as formas da Táboa 28 están presentes no TILG e/ou no CORGA, e todas elas, agás *buche*, *podrir* e *terreiro*, ocorren 20 ou máis veces cando menos nun dos dous corpus.

Ao lado das voces galegas cuxa presenza sería en principio explicable como solucións tradicionais que non tiveron entrada no galego estándar, existe outro grupo de exemplos, recollidos na Táboa 29, que son cando menos relativamente habituais no galego elaborado, pero non no galego popular, e que teñen orixes de diverso tipo.

TÁBOA 29. Formas transferidas desde o galego elaborado

Forma non estándar	Forma estándar
<i>soer</i> (25)	<i>acostumar</i> (6), <i>adoitar</i> (17)
<i>voltar</i> (8)	<i>volver</i> (96)
<i>nen</i> (5)	<i>nin</i> (147)
<i>blogue</i> (1)	<i>blog</i> (6)
<i>clube</i> (1)	<i>club</i> (0)
<i>evoluír</i> (1)	<i>evolucionar</i> (53)
<i>nembargantes</i> (1)	<i>porén</i> (63), <i>non obstante</i> (43), <i>con todo</i> (7), <i>así e todo</i> (5), <i>no entanto</i> (1), <i>así a todo</i> (0)
<i>respostar</i> (1)	<i>responder</i> (10)

O verbo *soer*, que figura na Táboa 7 por ter máis ocorrencias ca *adoitar*, xa foi mencionado en §3.2.1.2, como unha posible adaptación desde o castelán *soler*, pero ademais indicamos que a súa presenza nos textos escritos (1103 exemplos no *TILG*, 516 no *CORGA*) tamén podía ser debida á recuperación que experimentou a forma galega arcaica (González Seoane, 2006, *Dicionario de dicionarios do galego medieval*, s.v. *soer*) no galego elaborado (Fernández Salgado, 2004, s.v. *soler*).

Con respecto a *voltar*, co significado de ‘volver’, é unha forma con moita tradición nos textos galegos escritos (5030 casos no *TILG*, 2954 no *CORGA*) que non se recolle no código normativo galego por non existir apenas no galego popular (rexístrase nun único punto da provincia de Ourense no *ALGa* e non se recolle no *TLPGP*). A súa orixe parece estar nunha transferencia desde o portugués, adaptada probablemente cun afán diferencial (véxase sobre esta forma Fernández Salgado, 2004, s.v.). A terceira forma máis frecuente é *nen*, voz tamén habitual no galego elaborado (2276 ocorrencias no *TILG*, 1199 no *CORGA*) e que pode resultar dunha revitalización do galego arcaico ou ben dunha transferencia do portugués. *Blogue* e *clube*, cun único exemplo de cada caso, son adaptacións de préstamos en que se produce a adición dun <e> despois dunha oclusiva en posición final, coma no portugués. Rexistramos 249 exemplos de *clube* no *TILG* e 492 no *CORGA*, mentres que, con respecto a *blogue*, palabra máis recente, hai 37 casos no *CORGA* e ningún no *TILG*. *Respostar*, co valor de ‘responder’, é tamén unha forma innovadora, presente en textos escritos (540 exemplos no *TILG*, 631 no *CORGA*), que probablemente resulte da influencia do verbo portugués coincidente formalmente, aínda que nesta lingua o significado é máis restrinxido (o *Dicionário Priberam da língua portuguesa*, s.v., define como ‘responder incivilmente’).

Evoluír é tamén unha forma tirada do portugués, que á súa vez a adapta do francés *évoluer* e que ten certa tradición na lingua escrita (35 casos no *TILG*, 70 no *CORGA*).

Finalmente, *nembargantes* é unha forma que estivo admitida polo código normativo galego ata o ano 2003 e que ofrece abondos testemuños na lingua escrita (590 no *TILG* e 610 no *CORGA*). Foi desbotada pola súa posible orixe no español *no embargante* (*vid.* Freixeiro Mato, 2005, p. 261), aínda que non a consideramos dentro das transferencias do español adaptadas tanto polas diferenzas formais entre a forma galega e a española como polo feito de que a súa presenza nos textos actuais parece dificilmente explicable por transferencia desde o castelán, onde a forma *no embargante* é moi pouco usual, tal e como indica o propio *Diccionario de la Real Academia Española* (s.v. *embargante*).

3.2.1.6. *Estranxeirismos non adaptados*

Neste grupo incluímos 7 palabras (4 distintas), etiquetadas con L_for, o que supón que o 0,23% dos códigos explicativos asignados ás palabras con diferenzas que van máis alá do xénero e da acentuación corresponden a esta orixe. Caracterízanse porque nelas figura un estranxeirismo na súa forma orixinal (ou próxima), cando no estándar se propón a súa adaptación. Son as palabras que se recollen na Táboa 30:

TÁBOA 30. Estranxeirismos non adaptados

Forma non estándar	Forma estándar
<i>tablet / táblet</i> (3)	<i>tableta</i> (0)
<i>gasoil / gasioi</i> (2)	<i>gasóleo</i> (0)
<i>amateur</i> (1)	<i>amador</i> (0)
<i>yankee</i> (1)	<i>ianqui</i> (0)

Trátase de catro substantivos, tres deles de orixe inglesa e un de orixe francesa. Tanto *amateur*, como *gasoil* e *tablet* teñen certa presenza nos textos galegos, tal e como se pode comprobar na Táboa 46 do Apéndice.

3.2.1.7. *Mestura ou confusión entre palabras ou expresións*

Inclúense neste grupo seis formas cunha única ocorrencia en cada caso. O código correspondente, L_mix, supón un 0,20% do total das etiquetas asignadas ás formas con diferenzas que van máis alá do xénero e da acentuación. Trátase de elementos que parecen combinar elementos de varias palabras ou expresións e son as que se recollen na Táboa 31.

TÁBOA 31. Formas resultantes da mestura de palabras e expresións

Forma non estándar	Formas combinadas
<i>a saber Deus</i> (1)	<i>a saber</i> (0) + <i>sabe Deus</i> (0)
<i>adedicar</i> (1)	<i>adicar</i> (44) + <i>dedicar</i> (71)
<i>dar lugar a cabo</i> (1)	<i>dar lugar</i> (26) + <i>levar a cabo</i> (33)
<i>indixesta</i> (subst.) (1)	<i>inxesta</i> (1) + <i>indixesto</i> (0)
<i>non entanto</i> (1)	<i>non obstante</i> (43)+ <i>no entanto</i> (1)
<i>remontar a vista</i> (1)	<i>remontarse</i> (0)+ <i>botar a vista</i> (0)

Como se pode observar, en dous casos atopamos a combinación de elementos de varias palabras, relacionadas semanticamente, para conformar unha nova unidade, o verbo *adedicar*²⁸ por un lado e o substantivo *indixesta* polo outro, no contexto «Adultos e nenos con graves problemas de saúde pola indixesta sen control de comida basura». Nesta última forma, onde o esperable sería *inxesta*, parece haber unha mestura desta voz, apropiada ao contexto, co adxectivo *indixesto* ('Que non se dixire ou que se dixire con dificultade'), relacionado tamén co ámbito da alimentación.

Nos outros exemplos, atopamos a combinación de diferentes expresións, sobre todo expresións pluriverbais, sinónimas ou en todo caso próximas semanticamente, para dar lugar a unha nova unidade. No caso de *non entanto* (recollida como *non en tanto*), o valor contraargumentativo do conector, e non só a forma *non obstante*, pode favorecer a aparición do elemento negativo na expresión. Canda *adedicar*, só esta forma, con 7 exemplos no *TILG* e 17 no *CORGA*, teñen presenza en corpus e ningunha figura no *TLPGP*.

A proximidade semántica das expresións que están na base da nova forma é esperable. Estas mesturas téñense estudado no ámbito da fraseoloxía e, de acordo coas investigacións sobre os «idioms blends», «idioms with similar meaning are more likely to cause mix-ups than idioms of dissimilar meaning» (Karlsson, 2019, p. 182).

3.2.1.8. Formas de difícil clasificación

Un total de catro formas con diferenzas con relación ao estándar que van máis alá do xénero e da acentuación non se agrupan en ningunha das clases establecidas e non reciben, por tanto, ningún código no campo que indica

²⁸ Como se indica na Táboa 47 do Apéndice, no *TILG* recóllense cinco exemplos do verbo *adedicar*, que se poderían explicar tamén a través da presenza dun a- protético. Con todo, o verbo *adedicar* non se recolle nin no *TLPGP* nin no *GVGH* (no *ALGa* non hai preguntas que xeren esta resposta ou variantes), co cal non consideramos esta forma como variante galega non estándar (L_gal).

a orixe da diverxencia con respecto ao estándar. Son as que se recollen na Táboa 32:

TÁBOA 32. Formas non estándar con orixe de difícil clasificación

Forma non estándar	Forma estándar
<i>estrea</i> (1)	<i>estrela</i> (39)
<i>fugaxidade</i> (1)	<i>fugacidade</i> (1)
<i>lumen</i> (1)	<i>lume</i> (1)
<i>rexeitar</i> (1)	<i>rexeitar</i> (11)

É moi probable que cando menos algunhas de tales formas sexan simples lapsos de escritura, pero optamos por consideralas aquí porque tamén poden ser o resultado dun mal coñecemento da forma estándar.

En relación con *estrea*, debemos lembrar a forma *estreia* recollida en §3.2.1.2, que mostra a existencia dalgunhas dificultades con esta voz, e en *fugaxidade* resulta complicado explicar a presenza do <x> como lapso, pois este son non aparece no seu contexto próximo. No caso de *lumen*, non podemos deixar de mencionar a coincidencia coa forma latina, aínda que a presenza previa da palabra *descubrimiento*, coa mesma sílaba <men> que atopamos en *lumen* («descubrimiento do lumen») pode tamén explicar, como lapso, o <n> final. Por outro lado, *rexeitar*, coa presenza do dobre ditongo, é un bo candidato para ser considerado un lapso, pero tamén é certo que figuran varios exemplos desta palabra na rede e mesmo nalgún caso con máis dunha ocorrencia na mesma páxina²⁹, así como dous exemplos no *CORGA* (trátase da única das catro formas deste grupo rexistrada nalgunha das tres fontes presentadas no Apéndice).

3.2.2. Formas con diverxencias de xénero

Tal e como se indicou en §2, establecemos unha clase específica, etiquetada con L_gen_su, con aquelas formas que presentan un xénero diferente do estándar. Antes de amosar os resultados, cómpre delimitar adecuadamente esta clase de diverxencias, diferenciándoas doutras desviacións que afectan ao xénero, pero que se sitúan no nivel gramatical, codificadas en *CORTEGAL* con G_gen_su. Tal e como indican Álvarez de la Granja e Rodríguez Rodríguez (neste volume, §2.5.6), esta última etiqueta emprégase para casos en

²⁹ Por exemplo en <http://www.galiciaconfidencial.com/noticia/57982-compromiso-denuncia-insostible-inaceptable-posicion-marea-bng-resolucion-venezuela> [última consulta 04/03/2023], onde atopamos *rexeitamento* e *rexeitar*.

que se produce unha escolla inadecuada do xénero gramatical dunha palabra variable, de modo que se xera un problema de concordancia (véxase o capítulo de Cidrás neste volume). Velaquí algúns exemplos, onde destacamos en cursiva as formas que non concordan cos elementos aos que remiten, modifican, determinan...:

- (6) As persoas somos *consumidores* dende *pequenos* ata a morte
- (7) Calquer cousa que sexa extrema penso que non é *bó*
- (8) *as* prezos tenderán a subir
- (9) creo que iso lle dou visibilidade a un sector que non estaba moi recoñecido e fixo que a sociedade comezara a interesarse por *ela*

Hipoteticamente, sería posible considerar que nos exemplos anteriores se lles asigna un xénero inadecuado a *persoas*, *cousa*, *prezos* e *sector*, pero, por un lado, isto queda descartado polo feito de que outras palabras dos fragmentos mostran a concordancia adecuada («as [persoas]», «extrema», «un [sector]», «recoñecido»), porque no mesmo texto podemos encontrar concordancias correctas («polos prezos» no texto de onde se extraeu o exemplo 8) ou simplemente polo feito de que, como xa indicamos, o alumnado que escribe os textos de *CORTEGAL*, de maneira xeral, estuda galego e recibe clases en galego cando menos desde a educación primaria, de xeito que resultaría realmente chocante que se realizase unha atribución de xénero inadecuada ás palabras mencionadas. Así pois, consideramos tales casos como problemas de concordancia, que se poden explicar por diferentes motivos, como a distracción, a reformulación do que se pensaba dicir inicialmente ou a concordancia *ad sensum* (*vid.* o capítulo de Cidrás neste volume).

No entanto, unha vez máis existen algúns casos en que si parece plausible considerar que a/o estudante realiza unha atribución inadecuada de xénero. É o que sucede en (10), (11) ou (12):

- (10) Persoalmente, considero que esta nova cociña é *un arte*, *un arte* que como *todos* estase a converter nun negocio moi rentable
- (11) Cada un ten *as súas* costumes *gastronómicas*
- (12) As empresas teñen *unha* gran labor, *a* de crear moitos e moi rapido

Como se pode observar nos exemplos presentados, a/o estudante parece atribuírle a *arte* xénero masculino, e a *costume* e a *labor* feminino, o que se reflicte en varios dos elementos que os acompañan, destacados en cursiva. Sendo isto así, non hai nestes casos ningún problema de concordancia, posto que a/o estudante concorda adecuadamente de acordo co xénero que lles asigna aos substantivos indicados, senón que tal atribución non se

corresponde coa propia do código normativo do galego. En tales casos, temos un problema de carácter léxico, na medida en que «a propiedade de xénero no substantivo non se recibe por ningunha regra, senón que forma parte da entrada léxica» (Dubert, 2010, p. 64).

Agora ben, para considerar que existe un problema léxico de asignación inadecuada de xénero debemos apoiarnos nalgún argumento que sustente esta interpretación, como a recorrencia do mesmo problema no corpus ou noutros textos galegos ou a existencia de fenómenos de analoxía (inter ou intralingüística) que poidan explicar a atribución inadecuada. En calquera caso, tal e como se indica en Álvarez de la Granja e Rodríguez Rodríguez neste volume (§2.5.6), a clasificación e anotación ten carácter subxectivo, maior ou menor segundo os casos, e, a pesar do emprego dos indicios sinalados, este é un exemplo claro desta subxectivade, pois a consideración no nivel gramatical ou no nivel léxico das desviacións do estándar que estamos comentando responde en última instancia a unha interpretación da persoa que analiza os textos.

As discrepancias entre o xénero da palabra galega e o do seu cognado en castelán explican a maior parte das 138 formas clasificadas neste grupo, correspondentes a 26 substantivos distintos, tal e como se pode comprobar na Táboa 33, onde figuran os casos con probable influencia do español (aquí e nas seguintes táboas «m» significa masculino e «f» feminino):

TÁBOA 33. Formas con asignación inadecuada de xénero por influencia do castelán

Xénero non estándar	Número de casos	Xénero estándar	Número de casos	Porcentaxe de presenza do xénero non estándar
<i>arte</i> (m)	35	<i>arte</i> (f)	49	41,67%
<i>costume</i> (f)	22	<i>costume</i> (m)	11	66,66%
<i>labor</i> (f)	15	<i>labor</i> (m)	9	62,5 %
<i>leite</i> (f)	15	<i>leite</i> (m)	30	33,33%
<i>canle</i> (m)	5	<i>canle</i> (f)	10	33,33%
<i>orixe</i> (m)	4	<i>orixe</i> (f)	8	33,33%
<i>viaxe</i> (m)	4	<i>viaxe</i> (f)	12	25,00%
<i>porcentaxe</i> (m)	3	<i>porcentaxe</i> (f)	4	42,86%
<i>análise</i> (m)	2	<i>análise</i> (f)	1	66,67%
<i>engranaxe</i> (m)	2	<i>engranaxe</i> (f)	0	100%
<i>equipaxe</i> (m)	2	<i>equipaxe</i> (f)	0	100%
<i>marxe</i> (m)	2	<i>marxe</i> (f)	2	50%

TÁBOA 32. Continuación

<i>aprendizaxe</i> (m)	1	<i>aprendizaxe</i> (f)	2	33,33%
<i>arma</i> (m)	1	<i>arma</i> (f)	1	50%
<i>color</i> (m)	1	<i>color</i> (f)	0	100%
<i>emplataxe</i> (m)	1	<i>emplataxe</i> (f)	0	100%
<i>linguaxe</i> (m)	1	<i>linguaxe</i> (f)	0	100%
<i>paisaxe</i> (m)	1	<i>paisaxe</i> (f)	1	50%
<i>sal</i> (f)	1	<i>sal</i> (m)	0	100%
<i>sangue</i> (f)	1	<i>sangue</i> (m)	1	50%
Total	119		141	

Como se pode observar, son 20 formas diferentes caracterizadas, na súa gran maioría (agás *arma*, que se comenta embaixo), porque o cognado español ten un xénero diferente da voz galega estándar, o que probablemente explicaría a presenza do xénero non estándar en *CORTEGAL*. Deixando a un lado as voces con moi escasos rexistros, a forma con xénero estándar adoita ser máis frecuente ca a palabra co xénero do español, aínda que *costume* (f) e *labor* (f) supoñen unha excepción. En calquera caso, debe terse en conta que na Táboa 33 e nas seguintes deste apartado só computamos as formas que aparecen en contextos que permiten determinar o xénero atribuído, quedando fóra do cómputo aqueles usos en que resulta imposible asignar xénero gramatical (por exemplo, «Para gustos, sempre hai *colores*»).

Por outro lado, cómpre sinalar que non se consideran como desviacións léxicas dous casos en que a voz comeza con <a> tónico e o artigo que precede a forma coincide co masculino («o alza» e «un aula»), non habendo ningunha outra forma do texto que permita coñecer o xénero atribuído ao substantivo. Consideramos que tales casos se poden entender como aplicacións da norma do español (inexistente en galego) que establece que diante de substantivos femininos que comezan por <a> tónico o artigo definido que cómpre usar é *el* e o artigo indefinido máis frecuente e recomendado é *un* (Real Academia Española, 2005, s.v. *uno -na*). Dado que este emprego non implica cambio de xénero, non o interpretamos como un problema léxico, senón como un problema gramatical de selección inadecuada de determinante (que se codifica con G_det_su). No caso de *arma*, que si aparece na Táboa 33, o contexto dá a entender que o xénero atribuído é o masculino («un bo arma»), o que pode explicarse por unha reinterpretación en español, que se transmitiría ao galego, do substantivo como masculino, derivada do emprego

dos determinantes arriba citados (desta reinterpretación temos varios rexistros en castelán, como por exemplo «podría ser un buen arma contra este tipo de gases nerviosos», *Corpus del español: Web/Dialects*, Davies, 2016).

Finalmente, no caso de *arte*, que aparece claramente como masculino en numerosos contextos (vid, a modo de ilustración. o exemplo 10 ou estoutro: «pasión polo arte culinario»), optamos por computar como xénero non estándar tamén os casos en que o contexto non nos permite saber se se está atribuíndo o xénero masculino ou se se está aplicando a regra de selección do determinante arriba indicada (por exemplo, «a gastronomía é un arte máis nos nosos días»). Nótese que, desta volta, a diferenza de *alza* e *aula*, a voz ten en castelán ambos os xéneros e en singular é máis frecuente en masculino (Real Academia Española, 2005, s.v. *arte*), o que pode dar lugar á atribución dun xénero inadecuado por transferencia desde a lingua veciña.

Noutro grupo de exemplos o xénero atribuído ás voces pode explicarse por un proceso de analogía realizado dentro do propio galego. Os casos ofrécese na Táboa 34, onde «c» significa xénero común:

TÁBOA 34. Formas con asignación inadecuada de xénero por analogía

Xénero non estándar	Número de casos	Xénero estándar	Número de casos	Porcentaxe de presenza do xénero non estándar
<i>interese</i> (f)	3	<i>interese</i> (m)	74	3,90%
<i>auxe</i> (f)	2	<i>auxe</i> (m)	86	2,27%
<i>personaxe</i> / <i>persoaxe</i> (f)	3	<i>personaxe</i> / <i>persoaxe</i> (c)	9	25%
Total	8		169	

Trátase de dúas formas, *interese* e *auxe*, cuxa presenza no corpus co xénero non estándar é moi pouco significativa, e das dúas variantes *personaxe* e *persoaxe*, cunha porcentaxe de aparición maior, pero que, en todo caso, non supera o 25%. No caso de *interese*, o emprego do feminino pode explicarse probablemente porque a maior parte das palabras rematadas en <-ese> en galego son femininas (só figuran tres exemplos de formas masculinas no DRAG, *interese*, *desinterese* e *ese*, fronte a 86 femininas), o que conduciría á asignación a *interese*, por analogía, deste último xénero.

De xeito parello, a maior parte dos substantivos acabados en <-xe> en galego son palabras femininas (27 substantivos masculinos, fronte a 194 femininos no DRAG, sobre todo polo gran número de voces formadas co sufixo *-axe*, que selecciona feminino para a palabra resultante). Este pode ser o

motivo de que, por analogía, se lle asigne xénero feminino á palabra masculina *auxe*, así como a *personaxe / persoaxe* («o que importa é o fútbol e a vida das *distintas personaxes públicas*»), que é unha voz de xénero común na variedade estándar. Neste caso, ademais, produciuse un cambio na proposta normativa, que inicialmente presentaba esta forma como un substantivo feminino invariable, o que tamén pode explicar a frecuencia con que se atopa o seu emprego en feminino en textos galegos para designar individuos de sexo masculino ou para usos xenéricos.

Finalmente, presentamos outros dous exemplos, recollidos na Táboa 35, en que a desviación de xénero se explica por motivos diferentes aos indicados.

TÁBOA 35. Formas cuxa asignación inadecuada de xénero non se explica nin por transferencia nin por analogía

Xénero non estándar	Número de casos	Xénero estándar	Número de casos	Porcentaxe de presenza do xénero non estándar
<i>sabor</i> (f)	2	<i>sabor</i> (m)	87	2,25%
<i>Internet</i> (m)	9	<i>Internet</i> (f)	5	64,29%
Total	11		92	

Encontramos por unha banda dous casos en dous textos distintos de emprego de *sabor* como palabra feminina (2,25%), fronte a 87 usos como masculina: «unha comida con presentación e *unha sabor perfectas* non é sinxelo», «salsas que adornen *toda a sabor* da comida tradicional». É moi posible que estes usos respondan a un procedemento de hipercorrección. Aínda que a maior parte dos substantivos galegos que rematan en <-or> son masculinos, na listaxe de femininos «problemáticos» que se adoita estudar (aqueles que teñen un xénero diferente do cognado do castelán) inclúense sempre varias formas rematadas en -or (*calor, cor / color, dor / dolor, suor*), e é probable que as/os estudantes ampliasen equivocadamente estes casos de discrepancia entre galego e castelán á forma *sabor*.

Finalmente, encontramos 9 usos de *Internet* en xénero masculino (64,29%), fronte a 5 en feminino, que é o xénero estándar (na maioría dos usos non e posible identificar o xénero). Debe terse en conta que, malia existir unha proposta normativa, estamos ante un neoloxismo que non fixou aínda o seu xénero definitivamente na fala, como mostra da inestabilidade característica das unidades de recente formación (Cabré, 2003, p. 30).

Practicamente todos os casos explicables por transferencia do xénero do cognado español están presentes en corpus, con máis de 20 ocorrencias cando

menos nun deles. As únicas excepcións son as formas *engranaxe* e *emplataxe*. Tamén son frecuentes en corpus *personaxe*, como forma invariable de xénero feminino, e en menor medida *Internet* (f). No galego popular do TLPGP só se recollen con xénero non estándar formas cuxo xénero é explicable por influencia do castelán: *arte*, *canle*, *color*, *labor*, *leite*, *viaxe* e *sangue*.

3.2.3. Formas con diverxencias de acentuación

Tal e como se indicou en §2, constitúen un grupo específico aquelas formas, etiquetadas con L_ac_su, que portan o acento nunha sílaba diferente da que corresponde na forma estándar. Estas desviacións deben distinguirse das que se sitúan no nivel ortográfico, omisión de acento gráfico (O_ac_om), adición de acento gráfico (O_ac_ad) e cambio de lugar do acento gráfico (O_ac_wp), estudadas en López-Sández e Lorenzo-Herrera neste volume. Neste último caso non presupoñemos un cambio de lugar do acento de intensidade, senón un erro na asignación da marca gráfica, que pode ocorrer por descoñecemento das regras, por descoido, pola rapidez con que se elabora a tarefa etc. Encontramos aquí outro bo exemplo do carácter subxectivo da anotación, pois a consideración dunha desviación no nivel ortográfico ou no nivel léxico responde ás claras á interpretación da persoa que codifica os textos.

Deste modo, en formas recollidas en *CORTEGAL* como *facil* (por *fácil*) ou *máis* (por *máis*), poderíamos supoñer asociadas a tales formas gráficas as pronuncias, con cambio de lugar da sílaba tónica, [fa'θil] e [ma'is]. No entanto, debe terse en conta que, como xa indicamos en varias ocasións, o alumnado que escribe os textos do corpus, na súa gran maioría, estuda galego e recibe clases en galego cando menos desde a educación primaria, e sabemos tamén que o estudantado ten especiais problemas á hora da acentuación gráfica en galego (con diferenzas significativas en relación co castelán, de acordo cos datos de Loredó e Silva Valdivia, 2020, p. 70). Neste marco, parece lóxico pensar que en palabras (tan frecuentes) como as indicadas, estamos ante un problema meramente ortográfico e non ante a modificación da posición da sílaba tónica da palabra, aínda que, por suposto, nunca teremos certeza plena ao respecto.

Agora ben, existen unhas poucas formas en que se pode formular máis plausiblemente a hipótese de que a alumna ou o alumno situou a sílaba tónica nun lugar que non lle corresponde (aplicando correctamente as regras de acentuación gráfica de acordo coa posición outorgada ao acento de intensidade e non cometendo, daquela, ningún erro ortográfico, senón léxico). Os exemplos máis evidentes son aquelas formas que presentan a acentuación coincidente coa do cognado español. Podemos citar, por exemplo, a forma

élite, con acentuación esdrúxula, que aparece sete veces no corpus, fronte aos dous casos da forma estándar *elites*, con acentuación grave. Non é por casualidade que a forma *élite* adoite aparecer nas listaxes de formas non estándar que se deben evitar (ten entrada, por exemplo, en Fernández Salgado, 2004), senón que este feito responde á súa presenza frecuente nos textos (por exemplo, «e con esas motoserras do trinque e eses | (INH) uniformes tipo corpo de *élite* | parecen as ss hitlerianas do alcalde», fragmento extraído do corpus de textos orais *CORILGA*, Regueira, 2012). É este tipo de formas as que consideramos neste apartado. Son un total de 24 casos e 8 formas distintas. Os exemplos con probable influencia do español son os que figuran na Táboa 36.

TÁBOA 36. Formas con cambio na posición do acento de intensidade por probable influencia do español

Acentuación non estándar	Número de casos	Acentuación estándar	Número de casos	Porcentaxe de presenza da acentuación non estándar
<i>élite</i>	7	<i>elite</i>	2	77,78%
<i>textil</i>	5	<i>téxtil</i>	14	26,32%
<i>coltán</i>	2	<i>coltan</i>	0	100%
<i>periodo</i>	3	<i>período</i>	14	17,65%
<i>vértixe</i>	1	<i>vertixe</i>	0	100%

Evidentemente, e sobre todo nas formas en que desaparece o acento gráfico (*periodo*, *textil*), dado que os problemas ortográficos de omisión da marca gráfica son moito máis frecuentes ca os de adición (*vid.* a Táboa 9 en Álvarez de la Granja e Rodríguez Rodríguez e §2 en López-Sández e Lorenzo-Herrera, ambos nesta obra), é tamén posible que esteamos ante erros meramente gráficos³⁰.

Ademais das formas con presumible influencia do español, temos en *CORTEGAL* outros exemplos con acento de intensidade non estándar, recollidos na Táboa 37.

³⁰ No caso de *textil* debemos sinalar que entendemos os dous casos do plural *textiles* que figuran no corpus como erros ortográficos de omisión de acento e non como erros léxicos, posto que o plural estándar da forma normativa *téxtil* é *téxtiles* e a de *textil* sería *textís*. No entanto, non podemos esquecer que existen no corpus varios casos de emprego do plural en *-es* das formas agudas rematadas en *-l* (*perfiles*, *materiales*, *locales...*), de tal xeito que é moi posible que eses plurais sexan tamén en realidade dous casos de desprazamento da sílaba tónica acompañados dunha formación non estándar do plural.

TÁBOA 37. Formas con cambio na posición do acento de intensidade non explicables por influencia do español

Acentuación non estándar	Número de casos	Acentuación estándar	Número de casos	Porcentaxe de presenza da forma non estándar
<i>canón</i>	4	<i>canon</i>	14	22,22%
<i>Nóbel</i>	1	<i>Nobel</i>	0	100%
<i>nóvel</i>	1	<i>novel</i>	0	100%

Aparecen catro exemplos de *canón*, dous deles repetidos no mesmo texto, o que é un indicio de que se trata dun cambio de sílaba tónica. Cómpre indicar que a forma *canón* sería homógrafa da voz empregada para a peza de artilleería, e que é posible que unha confusión con esta forma estea na orixe do problema detectado. Debe facerse constar, finalmente, que a presenza relativamente elevada da voz *canon* no corpus responde ao feito de ser mencionada (en singular: *canon*) no texto obxecto de comentario³¹.

Con respecto a *Nóbel*, tal e como se indica na Táboa 50 do Apéndice, rexistramos 18 exemplos no *CORGA* que presentan esta acentuación gráfica non estándar. Trátase, pois, dunha acentuación que non é infrecuente e que pode ter influencia na pronuncia do adxectivo, rexistrada unha vez tamén no *CORGA* («artistas asturianos consagrados e nóveles»). Nótese, neste sentido, que ás veces se confunden graficamente as dúas formas («un novel de literatura», *CORGA*).

4. DISCUSIÓN E CONCLUSIÓNS

Tal e como indicamos en §3.1, o 1,24% das palabras ou expresións do corpus son unidades léxicas non estándar, deixando a un lado problemas ortográficos, morfolóxicos, semánticos e discursivos. Non parece a primeira vista unha cifra especialmente elevada, pero non podemos esquecer que, como calquera corpus, *CORTEGAL* inclúe unhas poucas palabras altamente recorrentes, nas que a presenza dunha forma non estándar é pouco previsible. Así, por exemplo, simplemente os artigos, definido e indefinido, canda as preposicións *a*, *de* e *en*, supoñen o 25% das voces do corpus.

³¹ No caso do plural, encontramos catro exemplos de *canones* (tres deles recollidos nun mesmo texto), ao lado doutros catro de *cánones* e dun da forma estándar *canons*. Aínda que os catro exemplos de *canones* se poderían considerar como desviacións léxicas por cambio de posición do acento de intensidade, optamos por consideralos erros ortográficos de omisión do acento, posto que o plural en *-es* das formas agudas acabadas en <-n> é moi escaso no corpus (só temos sete exemplos: *aficiones*, *fogones*, *razones*...). Porén, o plural das poucas formas graves acabadas en *-n* existentes parece ofrecer máis problemas, como mostra a presenza dos catro exemplos de *cánones* arriba mencionados. Así pois, a formación do plural en *-es* na forma *canones* probablemente deba vincularse coa pronuncia da palabra como grave.

Se tomamos como referencia os textos, a media de palabras non estándar nas redaccións de *CORTEGAL* (cunha extensión recomendada, lembremos, de entre 200 e 250 palabras) é de 3,03 por texto. Por outro lado, case o 90% deles inclúe algunha unidade léxica non estándar e practicamente o 21% teñen 5 ou máis formas deste tipo, o que deixa patente cando menos certas dificultades neste nivel lingüístico en relación co coñecemento do galego estándar. Non é menos certo, en calquera caso, que a maior parte das formas non estándar atopadas presentan ocorrencias únicas.

Á hora de interpretar os resultados atopados en *CORTEGAL*, non podemos esquecer que a tarefa que ten o alumnado por diante é a de elaborar un texto na variedade estándar do galego, empregando o código normativo, e que os desvíos desta variedade e deste código serán previsiblemente sancionados. Así pois, as diverxencias con respecto ao estándar non se poden explicar por licenzas expresivas, violación da tarefa encomendada, diferencialismo..., senón por descoñecemento, inseguridade, dificultades de memoria ou por un despiste na selección de lingua, os que denominamos lapsos bilingües. Agora ben, existen certas limitacións e certos aspectos que cómpre considerar no momento de empregar *CORTEGAL* para avaliar a competencia léxica do alumnado na variedade estándar do galego.

Por unha banda, a ausencia doutros traballos de características similares a *CORTEGAL* dificulta medir a significancia das cifras arriba indicadas. En calquera caso, as carencias no coñecemento do repertorio léxico estándar do galego xa foron postas de manifesto, para 4º da ESO, en Loredó e Silva Valdivia (2020, p. 143), onde a competencia léxica global para o alumnado deste curso se cualifica de baixa, cunha puntuación de 38,29 sobre 100 (en castelán 50,33). Con todo, a proba tipo test realizada neste estudo para medir a competencia léxica é dificilmente comparable cos textos de *CORTEGAL*. Podemos acudir entón á proba de expresión escrita do mesmo traballo, de características moi similares á que está na orixe dos textos de *CORTEGAL*. Nesta última, a cualificación obtida no apartado de «Precisión e corrección» (que tamén incluíría problemas que en *CORTEGAL* son considerados semánticos) foi de 63,5 sobre 100, fronte a 71,3 na mesma proba en castelán, unha valoración que pon de manifesto tamén algúns problemas no ámbito léxico e un menor dominio en galego ca en castelán.

Sería interesante contar cun recurso do mesmo tipo ca *CORTEGAL* con textos extraídos da proba de lingua castelá e literatura dos exames ABAU

(que inclúe tamén unha pregunta de comentario), para poder comparar os resultados, e particularmente o nivel de competencia léxica en ambas as dúas linguas á hora de acceder á universidade. É máis que probable que as transferencias do galego no castelán nese hipotético corpus fosen moito menos numerosas ca as transferencias do castelán no galego rexistradas no noso, o que probablemente determinaría un número de formas non estándar bastante máis reducido nos textos en español, dada a significancia da transferencia nos textos de *CORTEGAL*.

Esta afirmación baséase tanto na vinculación existente entre prestixio e transferencia, pola cal «prestige predicts the directionality of borrowing, from a more prestigious language to a less prestigious one» (Carling et al., 2019, p. 1), como nos resultados doutros traballos previos realizados con alumnado do mesmo nivel educativo. Así, ao comparar o léxico dispoñible do castelán e do galego, proporcionado por alumnado de segundo de Bacharelato de Galicia (López Meirama, 2008; López Meirama e Álvarez de la Granja, 2014), compróbase como o número de formas non estándar que resultan dun proceso de transferencia do castelán no léxico dispoñible do galego é moitísimo mais elevado ca o número de transferencias do galego no léxico dispoñible do español de Galicia. Acudindo a un centro de interese, «O corpo humano», analizado nos dous proxectos, atópanse 43 transferencias do galego no castelán no léxico dispoñible do español de Galicia nun total de 23430 respostas, o que supón un 0,18% do total destas (Álvarez de la Granja, 2011, p. 80), mentres que no caso do léxico dispoñible do galego o número de transferencias nese mesmo centro de interese é de 1796, sobre un total de 19753 respostas, o que supón unha porcentaxe do 9,09% (Álvarez de la Granja e López Meirama, 2021, p. 117).

Por outra banda, aínda que os textos de *CORTEGAL* son textos escritos, que poden ser revisados e monitorizados a partir dos coñecementos adquiridos sobre a variedade estándar, non podemos deixar de mencionar as especiais circunstancias en que se desenvolve a tarefa encomendada: a limitación temporal, a necesidade de responder a varias preguntas e a tensión e o nerviosismo derivados da relevancia que as probas ABAU teñen de cara ao acceso á universidade determinan que non se creen as circunstancias máis adecuadas para que a/o estudante realice a mellor das súas producións (González Álvarez, 1999, p. 209). Este feito, canda as inseguridades arriba mencionadas e a escasa exposición de parte do estudantado ao galego, pode explicar a presenza nun mesmo texto de formas estándar e non estándar, de voces galegas e do castelán, ou mesmo de distintas variantes non estándar.

Por exemplo, unha mesma persoa escribe «*calquera* figura» e liñas despois «*calquer* rama», outras dúas inclúen *año* e *ano*, ou *su* e *seu*, mentres que noutro texto alternan as formas non estándar *desvantaxa* e *desventaxa*.

Outra limitación á hora de empregar os resultados do corpus vén dada pola restrición temática dos textos. Estes, tal e como se indica en Álvarez de la Granja e Rodríguez Rodríguez neste volume (§2.1), abordan catro temas diferentes, de tal xeito que moito vocabulario non aparece, ou aparece en moi escasa medida, porque non é pertinente tematicamente. En consecuencia, resulta imposible extraer conclusións sobre o grao de coñecemento deste léxico ausente ou moi limitado e, máis en particular, elaborar unha listaxe completa de formas non estándar de ampla implantación na escrita do noso estudantado. Por suposto, e tal e como indican nesa mesma epígrafe de Álvarez de la Granja e Rodríguez Rodríguez, outra limitación deriva da imposibilidade de traballar con variables sociolingüísticas dos textos.

Malia o indicado nos parágrafos precedentes, *CORTEGAL* ofrece datos globais de grande interese sobre o nivel de uso do léxico non estándar e tamén permite extraer algunhas conclusións particulares sobre certas formas, que probablemente requiran dun reforzo particular nas aulas (*vid.*, por exemplo, a Táboa 7). Tamén nos permite tirar conclusións sobre a orixe dese léxico. Por unha banda, unha parte do vocabulario non estándar ten a súa fonte noutras linguas ou variedades lingüísticas: encontramos voces dialectais, como *caír* ou *bon*; palabras empregadas no galego culto, como *voltar*, *nembargantes* ou *soer*; voces coa forma da lingua de orixe, como *yankee* ou *tablet*, pero sobre todo, como xa indicamos, numerosas voces coincidentes co castelán, que nos textos de *CORTEGAL* poden ter a súa fonte directamente no español ou ben, no caso dos galegofalantes, seren castelanismos tirados do galego popular, ás veces adaptados. Por exemplo, *postre*, *cumpleaños*, *olvidar*, *salir*, *parexa...* son voces habituais nesta variedade lingüística. No caso das voces transferidas directamente desde o español podemos atopar simples despistes ou lapsos bilingües (*su*, *dos...*) ou formas que, equivocadamente, se postulan como idénticas nas dúas linguas (*pilar*, *rentable...*), dada a existencia de moitos equivalentes coincidentes. Por outra banda, tamén atopamos creacións das/dos estudantes, que fan hipóteses sobre como será a forma galega estándar a partir do establecemento de correlacións entre galego e castelán (*semexante*, *ganadeiro...*), mediante hipercorreccións que tamén se basean en correlacións entre esas dúas linguas (*vaciña*, *a sabor...*), a través de analoxías baseadas noutras formas do propio galego (*acadamento*, *consomer...*) ou mesmo mesturando palabras ou expresións (*indixesta*, *a saber Deus...*).

Tal e como tivemos ocasión de comprobar ao longo da exposición, e como se pode verificar no Apéndice, entre as formas non estándar rexistradas en *CORTEGAL* encontramos, ao carón de voces ben testemuñadas noutros textos orais ou escritos, bastantes palabras e expresións non recollidas ou con escasas ocorrencias noutros discursos. Atendendo ás 336 formas diferentes que figuran no Apéndice³², un 16,96% non se rexistra nin no *TILG*, nin no *CORGA* (e tampouco no *TLPGP*). Isto non implica necesariamente que todas elas sexan hápax (por exemplo, unha busca na rede ofrece varios exemplos de formas presentes en *CORTEGAL* non rexistradas nas fontes sinaladas, como *vaciña* ou *pisciña*), pero é no grupo de formas non rexistradas noutras fontes onde hai máis posibilidades de atopar formas non almacenadas no lexicón mental das/dos estudantes que as escriben. No sentido contrario, un 57,74% das formas do Apéndice aparecen rexistradas con 20 ou máis ocorrencias nalgún dos dous corpus considerados, e un 35,71% con 100 ou máis, e son estas as que teñen máis posibilidade de formar parte do repertorio léxico das/dos estudantes que as utilizaron. En calquera caso, a presenza en corpus non implica en absoluto a lexicalización, posto que se pode simplemente postular a activación das mesmas hipóteses e dos mesmos mecanismos creativos nas persoas autoras dos textos de *CORTEGAL* e nas persoas autoras dos textos do *CORGA* e do *TILG*.

No que respecta ao *TLPGP*, un 79,46% das palabras presentes no Apéndice non figuran neste recurso. Isto non quere dicir, tal e como xa indicamos, que todas elas estean ausentes do galego popular, polas restricións no tipo de léxico recollido normalmente nas obras que integran o *TLPGP*. Non debemos esquecer tampouco que en tal variedade son especialmente frecuentes as importacións directas do español, pouco representadas no Apéndice.

Por outro lado, o correlato estándar dalgunhas das formas non normativas rexistradas en *CORTEGAL* é relativamente ben coñecido (*porén, esquecer, prato...*), pero noutros casos, cómpre sinalalo, é alleo á maior parte dos galegos: que a voz *coltan* é grave en galego é un feito coñecido probablemente só por algúns especialistas. E dubidamos de que propostas normativas como *engrenaxe, padrón* (equivalente ao castelán *patrón*), *inmaturo* ou *subliminar* estean moi difundidas. De igual xeito, o carácter non estándar de voces moi frecuentes no galego culto como *plantexar, novidoso, soer, xunto con, en torno a, asequible...* é descoñecido por unha gran parte da poboación,

³² Lembremos que neste anexo non se recollen todas as formas non estándar rexistradas en *CORTEGAL*, posto que as transferencias directas, moi numerosas, quedan fóra na súa gran maioría. De entre elas, só figuran no Apéndice as recollidas na Táboa 7, caracterizadas por seren máis frecuentes ca as correspondentes voces estándar.

mesmo daquela máis instruída, e incluso por quen ten formación específica en lingua galega (cómpre confesar que no propio proceso de anotación do corpus aprendemos algunhas propostas normativas). Así pois, non podemos obviar o feito de que unha parte das carencias no coñecemento do léxico estándar non son específicas do alumnado galego, senón que existe un problema xeral de difusión do código normativo entre a poboación galega, propiciado en boa medida pola situación de marxinalidade do galego na sociedade e polo seu papel secundario no acceso ao coñecemento e á información.

En relación co que acabamos de indicar, a análise dos datos de *CORTEGAL* tamén nos suscita a reflexión sobre por que determinadas voces propostas para o léxico estándar (creadas, recuperadas ou revitalizadas) callaron máis doadamente ca outras: así, por exemplo, podemos confrontar o 90,38% de *saír* (9,62% de *salir*) co 46,81% de *esquecer* (53,19 % de *olvidar*), sendo tanto *saír* como *esquecer* formas desprazadas en boa parte de Galicia polos correspondentes castelanismos; ou podemos tamén contrastar o 50% de *reflectir* co 2,94% de *prexudicar*. A análise destas diferenzas, e a determinación das súas causas, algo que en todo caso queda fóra do alcance deste traballo, pode contribuír a encontrar mecanismos que axuden a difundir as formas menos implantadas.

Tal e como indicamos, a gran maioría das formas non estándar teñen a súa orixe, de maneira máis ou menos directa, no contacto co castelán, pois case un 90% das formas analizadas son transferencias do español, directas ou adaptadas, e a maior parte das formas etiquetadas como hipergaleguismos poden en realidade considerarse tamén como adaptacións desde o español. Por tal motivo, gustaríanos ofrecer algunhas reflexións sobre estas formas maioritarias en *CORTEGAL*. As palabras de Matras (2009, p. 74) sobre a orixe das transferencias lingüísticas explican moi ben os resultados atopados no corpus:

We are now in a position to attribute a new interpretation to the process known as interference or transfer: it is a process by which the speaker makes or attempts to make creative communicative use of elements of the combined, full repertoire of linguistic structures in a context that requires selection from just a subset of that repertoire, i.e. from the appropriate 'language'. In the case of learners or bilinguals with a dominant language, the trigger behind the process is often lack of proficiency in the selected language of the context, or else a lapse in memory (retrieval difficulties) or in control over the selection mechanism of word-forms and other structures.

A cita, aplicable ao alumnado de *CORTEGAL* castelanfalante, podería entenderse tamén ao alumnado galegofalante que ten o galego popular como

variedade inicial, substituíndo lingua («language») por variedade (de feito, Paradis 2004, p. 187 defende que a interferencia se produce da mesma maneira entre linguas diferentes ca entre distintos rexistros dunha mesma lingua). Trataríase desta volta da carencia de dominio dunha variedade lingüística por parte de falantes doutra variedade da mesma lingua. Lémbrese que entre ambas as variedades existen diferenzas significativas no ámbito léxico, ao presentar o galego popular un considerable número de transferencias do español, que procuran eliminarse no código normativo. Outro tipo de transferencias entre variantes, as das formas dialectais (etiquetadas con L_{gal}) son, como vimos, bastante escasas en comparación coas que se producen entre castelán e galego.

Se reparamos agora na categoría gramatical das formas transferidas, debemos referirnos á predominancia dos substantivos. Os datos ofrecidos na Táboa 4 para todas as formas léxicas non estándar son moi similares aos rexistrados unicamente para as transferencias directas e adaptadas do español, cun 43% de substantivos, fronte a pouco máis do 21% de representación desta categoría no total do corpus. Aínda que en *CORTEGAL*, tal e como indicamos, podemos atopar tanto importacións xa lexicalizadas como transferencias ocasionais e lapsos bilingües, non podemos deixar de mencionar a coincidencia destes datos cos atopados para outras linguas no ámbito dos préstamos, que amosan unha predominancia dos substantivos (Field, 2002, p. 36; Matras, 2009, p. 167; Tadmor, 2009, p. 61). Ao mesmo tempo, encontramos nas palabras gramaticais moitos menos casos de transferencia ca nas palabras léxicas, o cal tamén coincide cos datos encontrados para as transferencias lexicalizadas: «content words are more borrowable than function words» (Tadmor, 2009, p. 59).

Entre as palabras e expresións de *CORTEGAL* que supoñen unha transferencia desde o castelán son maioritarias as importacións directas (72,03% dos códigos de orixe) fronte ás adaptacións ou mesturas (17,97%). Estes resultados son esperables dada a proximidade das dúas linguas: por un lado, a porcentaxe de palabras graficamente idénticas en galego e castelán (*boca, libro, casa, comer...*) é moi elevada, o que favorece a hipótese da coincidencia e, con ela, a importación directa; por outro lado, unha gran parte das voces do castelán non coincidentes co galego estándar, como *venta, olvidar* ou *cercano*, as tres rexistradas en *CORTEGAL*, non presentan ningún trazo, fónico ou morfolóxico, que se poida considerar alleo ao galego. Deste xeito, para quen descoñeza a correspondente forma galega estándar, non hai ningún aspecto formal que poña en dúbida a «galeguidade» destas palabras.

No entanto, noutras ocasións o alumnado si identifica certos elementos lingüísticos como alleos ao galego, e, non coñecendo a forma estándar, realiza un calco ou unha adaptación de expresións e palabras do castelán. Neste último caso, o estudantado elabora hipóteses sobre cal será a forma normativa a partir da aplicación de padróns de equivalencia que resultan dunha comparativa entre galego e castelán, o que dá lugar a modificacións de palabras etiquetadas como castelás no seu repertorio lingüístico. Neste proceso, posibilitado pola proximidade das linguas, os elementos modificados poden ser as vogais (prodúcese sobre todo monotongación de ditongos), a acentuación gráfica, as consoantes (substitúese sobre todo a fricativa velar pola fricativa postalveolar) e particularmente as terminacións. Lémbrese, en calquera caso, que algunhas das formas «híbridas» presentadas en §3.2.1.2, poden ser formas xa lexicalizadas e almacenadas no repertorio lingüístico dos falantes, extraídas do galego popular ou do galego elaborado.

Gustaríanos deternos brevemente sobre estas formas híbridas, sexan calcos e palabras adaptadas. A súa existencia é unha boa mostra dun trazo que caracteriza as/os falantes: a súa creatividade. Tal e como sinala Yaron Matras na cita anteriormente reproducida, as transferencias implican en por si un uso creativo de elementos do repertorio lingüístico, pero se a isto sumamos a construción dunha nova forma, mediante tradución ou adaptación, a creatividade duplícase. As transferencias adaptadas do español (así como os hipergaleguismos, as formas creadas mediante analoxías con outras voces galegas ou mesmo as que resultan de mesturas entre palabras e expresións) poñen de relevo «the centre-stage position of the bilingual speaker as a creative communicator» (Matras, 2009, p. 6).

Ao mesmo tempo, estas formas híbridas, adaptadas do español, canda as transferencias directas, pero mesmo aínda en maior medida pola interacción entre as dúas linguas que subxace tras delas, son unha boa mostra da existencia dun repertorio multilingüe a disposición dos individuos no acto comunicativo: «multilingual speakers do not 'block' or 'switch off' one of their languages when communicating in another, but [...] they have the full, complex linguistic repertoire at their disposal at all times» (Matras, 2009, p. 5).

Deste xeito, a existencia destas formas híbridas casa moi ben coas palabras de Gugemberger (2008, p. 37) sobre a postura hibridista en lingüística, que concibe «el hablante como autor creativo que, al combinar diferentes elementos, crea formas y perspectivas nuevas respondiendo así a la dinámica lingüística en una sociedad bi- o multilingüe». Neste proceso creativo, o fenómeno da analoxía, sexa inter- ou intralingüística, ocupa un papel fundamental.

Por último, gustaríanos facer unha breve reflexión sobre a conveniencia de levar a cabo un tratamento integrado de galego e castelán na aula, de acordo coa estratexia metodolóxica coñecida como Tratamento Integrado de Linguas (*vid.*, por exemplo, Ruiz Pérez, 2008 ou Rama Lorenzo, 2020). Tal e como acabamos de indicar, ao longo deste capítulo comprobamos como o léxico do galego e o do castelán non se organiza en repertorios pechados, senón que existe unha continua interrelación entre unha e outra lingua (favorecida pola proximidade entre ambas), que, no caso de *CORTEGAL*, se manifesta sobre todo na transferencia directa de voces do castelán ao galego ou na súa adaptación de acordo co establecemento de padróns recorrentes entre ambos os dous idiomas. Sendo así, probablemente sería de interese que a aprendizaxe dunha e outra lingua tivese en consideración esa continua interacción, aproveitase o que ambas as dúas teñen en común e reparase nos aspectos diferenciais, pero evitando outorgarlle ao galego, como adoita ser habitual, un papel secundario e dependente do castelán. Desta maneira, tamén se podería aproveitar o ensino integrado do léxico para reflexionar sobre o diferente devir histórico das linguas (con base nas correlacións entre galego e castelán que o propio alumnado realiza), sobre o distinto uso que se pode facer dos procedementos morfolóxicos que teñen as linguas (gal. *troco*, pero esp. *trueque*; gal. *deterioración*, pero esp. *deterioro*) e sobre ese fondo común, compartido polas linguas románicas, máis alá do galego e do castelán, que pode ser unha chave para comprender e aprender novos idiomas e para coñecer máis sobre os propios (Escudé e Calvo del Olmo, 2019).

REFERENCIAS BIBLIOGRÁFICAS

- ALGA = GARCÍA, C. e SANTAMARINA, A. (1990-). *Atlas lingüístico galego*. Fundación Pedro Barrié de la Maza, Conde de Fenosa; Universidade de Santiago de Compostela.
- ÁLVAREZ, R. e XOVE, X. (2002). *Gramática da lingua galega*. Galaxia.
- ÁLVAREZ DE LA GRANJA, M. (2011). A presenza do galego no léxico dispoñible do español de Galicia. Análise formal e funcional. En B. López Meirama (Ed.), *Estudios sobre disponibilidad léxica en el español de Galicia* (pp. 17-102). Universidade de Santiago de Compostela.
- ÁLVAREZ DE LA GRANJA, M. e LÓPEZ MEIRAMA, B. (2021). La presencia del español en el léxico disponible del gallego. El centro de interés el cuerpo humano. En M. Serrano Zapata e M. Á. Calero Fernández (Eds.), *Aplicaciones de la disponibilidad léxica* (pp. 115-145). Tirant Humanidades.
- CABRÉ, T. (2003). A neoloxía, inevitable e necesaria para as linguas. En M. Álvarez de la Granja e E. González Seoane (Eds.), *A estandarización do léxico* (pp. 15-36). Instituto da Lingua Galega; Consello da Cultura Galega.

- CARLING G., CRONHAMN S., FARREN R., ALIYEV E. e FRID J. (2019) The causality of borrowing: Lexical loans in Eurasian languages. *PLoS ONE*, 14(10), artigo e0223588. <https://doi.org/10.1371/journal.pone.0223588>
- CORGA = Centro Ramón Piñeiro para a investigación en humanidades. *Corpus de Referencia do Galego Actual (CORGA)* (versión 4.0). <http://corpus.cirp.gal/corga/> [última consulta: 04/03/2023]
- DAVIES, M. (2016-). *Corpus del español: Web/Dialectos*. <https://www.corpusdelespanol.org/web-dial/> [última consulta: 04/03/2023]
- DAVIES, M. (2016-). *Corpus do portugués: Web/Dialectos*. <https://www.corpusdoportugues.org/web-dial/> [última consulta: 04/03/2023]
- DICIONARIO CASTELÁN-GALEGO. Rinoceronte Editora. <https://rinoceronte.gal/diccionario/> [última consulta: 04/03/2023]
- DICIONÁRIO PRIBERAM DA LÍNGUA PORTUGUESA (2008-2021). Priberam Informática. <https://diccionario.priberam.org> [última consulta: 04/03/2023]
- DIGALEGO. DICCIONARIO DE GALEGO (2013). Xunta de Galicia; Ir Indo. <https://digalego.xunta.gal/> [última consulta: 04/03/2023]
- DRAG = Real Academia Galega (2012). *Dicionario da Real Academia Galega*. Real Academia Galega. <https://academia.gal/diccionario> [última consulta: 04/03/2023]
- DUBERT GARCÍA, F. (2004). *Influente ou influínte?* Prescrición e uso da vogal temática verbal da C-III nos adxectivos en *-nte* na lingua galega. En R. Álvarez Blanco e H. Romero Monteagudo (Eds.), *Norma lingüística e variación. Unha perspectiva desde o idioma galego* (pp. 223-246). Consello da Cultura Galega; Instituto da Lingua Galega. http://consellodacultura.gal/mediateca/extras/2005_Norma-linguistica-e-variacion_Influente-ou-influente_Francisco-Dubert-Garcia_pp224-247.pdf
- DUBERT GARCÍA, F. (2005) Interferencias del castellano en el gallego popular. *Bulletin of Hispanic Studies*, 83(3), 271-291. <https://doi.org/10.3828/bhs.82.3.1>
- DUBERT GARCÍA, F. (2010). Sobre o lugar do xénero na gramática do galego. *Estudos de Lingüística Galega*, 2, 57-74.
- DUBERT GARCÍA, F. (2013). El yeísmo y el fortalecimiento de /j/ en Galicia. En R. Gómez e I. Molina Martos (Eds.), *Variación yeísta en el mundo hispánico* (pp. 39-58). Iberoamericana Vervuert. <https://doi.org/10.31819/9783954870691-003>
- ESCODÉ, P. e CALVO DEL OLMO, F. (2019). *Intercomprensão: a chave para as línguas*. Parábola.
- FERNÁNDEZ REI, F. (1990). *Dialectoloxía da lingua galega*. Xerais.
- FERNÁNDEZ SALGADO, B. (2004). *Dicionario Galaxia de usos e dificultades da lingua galega*. Galaxia.
- FERREIRO, M. (1999). *Gramática histórica galega. I. Fonética e morfosintaxe*. Laiovento.
- FIELD, F. W. (2002). *Linguistic borrowing in bilingual contexts*. John Benjamins. <https://doi.org/10.1075/slcs.62>
- FREIXEIRO MATO, X. R. (2005). *Non obstante, sen embargo, nembargante(s)* e outros adverbios conectores do galego escrito: perspectiva diacrónica e uso actual. *Verba*, 32, 247-274.

- GARCÍA, C. e GONZÁLEZ GONZÁLEZ, M. (Dirs.) (2005). *Dicionario castelán-galego da Real Academia Galega*. Fundación Pedro Barrié de la Maza, Conde de Fenosa; Real Academia Galega.
- GÓMEZ CAPUZ, J. (1998). *El préstamo lingüístico. Conceptos, problemas y métodos*. Universitat de València.
- GÓMEZ MOLINA, J. R. (1999-2000). Transferencia y cambio de código en una comunidad bilingüe. Área metropolitana de Valencia (I y II). *Contextos*, XVII/XVIII/33-36, 309-360.
- GONZÁLEZ ÁLVAREZ, E. (1999). Análisis de los errores léxico-semánticos. En L. 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* (pp. 207-270). Universidade de Santiago de Compostela.
- GONZÁLEZ GONZÁLEZ, M. (2002). *El gallego hablado de Cangas de Morrazo*. [Tese de doutoramento inédita]. Universidad de Oviedo.
- GONZÁLEZ SEOANE, E. (Coord.) (2006). *Dicionario de dicionarios do galego medieval*. Universidade de Santiago de Compostela; Universidade de Vigo. <http://ilg.usc.gal/ddgm/> [última consulta: 04/03/2023]
- GUGENBERGER, E. (2008). El castellano y las lenguas regionales en España: Bilingüismo e hibridación. En M. Doppelbauer e P. Cichon (Eds.), *La España multilingüe. Lenguas y políticas lingüísticas de España* (pp. 31-52). Praesens.
- KABATEK, J. (2000). *Os falantes como lingüistas. Tradición, innovación e interferencias no galego actual*. Xerais.
- KARLSSON, M. (2019). *Idiomatic Mastery in a First and Second Language*. Multilingual Matters. <https://doi.org/10.21832/9781788922371>
- KIESLER, R. (1993). La tipología de los préstamos lingüísticos: no sólo un problema de terminología. *Zeitschrift Für Romanische Philologie*, 109(5-6), 505-525. <https://doi.org/10.1515/zrph.1993.109.5-6.505>
- HASPELMATH, M. (2009). Lexical borrowing: Concepts and issues. En M. Haspelmath e U. Tadmor (Eds.), *Loanwords in the world's languages. A comparative handbook* (pp. 35-54). De Gruyter Mouton. <https://doi.org/10.1515/9783110218442.35>
- LÓPEZ MEIRAMA, B. (2008). *Léxico disponible en el español de Galicia*. Universidade de Santiago de Compostela.
- LÓPEZ MEIRAMA, B. e ÁLVAREZ DE LA GRANJA, M. (2014). *Léxico dispoñible do galego*. Universidade de Santiago de Compostela. <https://www.usc.gal/libros/gl/lingstica/125-lexico-disponible-do-galego-lexico-disponible-do-galego.html>
- LOREDO, X. e SILVA VALDIVIA, B. (Coords.) (2020). *Avaliación da competencia bilingüe nos idiomas galego e castelán do alumnado de 4º da ESO*. Real Academia Galega.
- MATRAS, Y. (2009). *Language contact*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809873>

- MONTEAGUDO, H. (2003). A demanda da norma. Avances, problemas e perspectivas no proceso de estandarización do idioma galego. En H. Monteagudo Romero e X. M. Bouzada Fernández (Coords.), *O proceso de normalización do idioma galego 1980-2000. III. Elaboración e difusión da lingua* (pp. 37-129). Consello da Cultura Galega. <https://publicacions.academia.gal/index.php/rag/catalog/view/252/248/152>
- PARADIS, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins. <https://doi.org/10.1075/sibil.18>
- RAMA LORENZO, S. (2020). O Tratamento Integrado de Linguas como instrumento para a normalización da lingua galega. *Revista galega de educación*, 78, 32-35.
- REAL ACADEMIA ESPAÑOLA. *Diccionario de la lengua española* (23.ª ed., versión 23.5). <https://dle.rae.es> [última consulta: 04/03/2023]
- REAL ACADEMIA ESPAÑOLA (2005). *Diccionario panhispánico de dudas* (versión en liña). <https://www.rae.es/dpd/> [última consulta: 04/03/2023]
- REAL ACADEMIA GALEGA (2012). *Normas ortográficas e morfolóxicas do idioma galego* (23.ª ed.). Real Academia Galega; Instituto da Lingua Galega.
- REGUEIRA, X. L. (Coord.) (2010). *Diccionario de pronuncia da lingua galega*. Real Academia Galega; Instituto da Lingua Galega.
- REGUEIRA, X. L. (Dir.) (2012-). *Corpus Oral Informatizado da Lingua Galega* (CORILGA). Universidade de Santiago de Compostela. <http://ilg.usc.gal/corilga/> [última consulta: 04/03/2023]
- RODRÍGUEZ Río, X. (1993). Os préstamos na produción lexicográfica, terminográfica e enciclopédica galega actual. En M. Álvarez de la Granja e E. González Seoane (Eds.), *A estandarización do léxico* (pp. 419-446). Instituto da Lingua Galega; Consello da Cultura Galega.
- RUIZ PÉREZ, T. (2008). El tratamiento integrado de lenguas. Construir una programación conjunta. *Textos. Didáctica de la lengua y la literatura*, 47, 46-58.
- SANTAMARINA, A. (Coord.) (2006). *Diccionario de diccionarios* (versión en liña). Universidade de Santiago de Compostela; Universidade de Vigo. <https://ilg.usc.gal/ddd/index.php>
- TADMOR, U. (2009). Loanwords in the world's languages: Finding and results. En M. Haspelmath e U. Tadmor (Eds.), *Loanwords in the world's languages. A comparative handbook* (pp. 55-75). De Gruyter Mouton. <https://doi.org/10.1515/9783110218442-55>
- TERGAL = TERMIGAL (Dir. M. González González) (2016-). *TERGAL: Banco de termos galegos recomendados* (versión 1.5.2). Centro Ramón Piñeiro para a Investigación en Humanidades. <http://www.cirp.gal/tergal> [última consulta: 04/03/2023]
- TILG = Santamarina, A. (Dir.). *Tesouro informatizado da lingua galega* [versión 4.1]. Instituto da Lingua Galega. <http://ilg.usc.gal/TILG/> [última consulta: 04/03/2023]
- TLPGP = Álvarez, R. (Coord.). *Tesouro do léxico patrimonial galego e portugués*. Instituto da Lingua Galega. <http://ilg.usc.gal/Tesouro> [última consulta: 04/03/2023]

APÉNDICE

TÁBOA 38. Presenza no *TILG*, *CORGA* e *TLPGP* das formas non estándar cun nivel de aparición en *CORTEGAL* igual ou superior ao das formas estándar correspondentes así como das súas variantes

	TILG	CORGA	TLPGP
<i>acarrear</i> (5)	31	33	4
<i>afición</i> (1)	4	7	0
<i>afición</i> (14)	216	55	0
<i>apariencia</i> (16)	128	192	0
<i>arreglar</i> (7)	324	55	13
<i>artiluxio</i> (5)	82	41	0
<i>asequible</i> (18)	18	77	0
<i>basura</i> (19)	2	30	1
<i>búsqueda</i> (11)	79	182	0
<i>cercano</i> (9)	197	318	0
<i>chamativo</i> (8)	78	206	0
<i>costar</i> (16)	778	314	8
<i>coste</i> (29)	106	136	0
<i>costo</i> (3)	36	35	0
<i>cumpleanos</i> (5)	8	12	2
<i>custe</i> (4)	109	260	0
<i>disfrutar</i> (92)	492	877	0
<i>disfrute</i> (8)	80	92	0
<i>esclavitude</i> (5)	7	7	0 ³³
<i>estropear</i> (5)	89	47	4
<i>goce</i> (2)	253	230	0
<i>grado</i> (8)	606	336	1
<i>hosteleiro</i> (12)	36	90	0
<i>hostelería</i> (30)	28	75	0
<i>hostelero</i> (1)	0	2	0
<i>inconveniente</i> (1)	23	5	0
<i>inconvente</i> (1)	0	0	0
<i>inconvinte</i> (8)	6	21	0

³³ Recóllese *esclavítu* en dous concellos.

TÁBOA 38. Continuación

	TILG	CORGA	TLPGP
<i>increíble</i> (6)	50	39	0
<i>inflúinte</i> (5)	30	97	0
<i>invertir</i> (20)	99	103	0
<i>llamativo</i> (1)	1	0	0
<i>mezcla</i> (11)	34	18	1
<i>mezclar</i> (9)	65	9	5
<i>mistura</i> (3)	323	131	5
<i>misturar</i> (2)	734	273	1
<i>olvidar</i> (25)	988	237	9
<i>padar</i> (1)	0	0	0
<i>paladar</i> (13)	107	36	6
<i>perxudicar</i> (33)	206	121	0
<i>perxudicial</i> (18)	92	53	0
<i>pilar</i> (14)	145	66	1
<i>postre</i> (9)	181	201	3
<i>preciado</i> (5)	78	59	0
<i>promocionar</i> (8)	60	561	0
<i>receta</i> (38)	218	34	1
<i>reflejar</i> (1)	0	0	0
<i>reflexar</i> (12)	507	514	0
<i>rentábel</i> (2)	6	26	0
<i>rentable</i> (12)	69	50	1
<i>rexir</i> (18)	436	104	2
<i>soer</i> (25)	1103	516	0
<i>soler</i> (3)	296	31	1
<i>tercermundista</i> (8)	0	1	0
<i>tontería</i> (7)	81	84	0
<i>vacío</i> (10)	329	115	2
<i>venta</i> (32)	250	122	0
<i>xunto con</i> (16)	1165	3336	0

TÁBOA 39. Presenza no *TILG*, *CORGA* e *TLPGP* de calcos estruturais do español

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
Calcos estruturais de expresións pluriverbais fixas do español			
<i>á longa</i> (2)	11	19	0
<i>comerse o mundo</i> (1) ³⁴	2	2	0
<i>dárseas</i> (1)	1	0	0
<i>de campionato</i> (1)	3	17	0
<i>hoxe día</i> (2)	1	7	0
<i>perrito quente</i> (1)	0	0	0
<i>(a) pescada que se morde a cola</i> (1)	1	3	0
<i>(unha) pescadilla que se morde / mórdese a cola</i> (2)	0	0	0
<i>picar o gusaniño</i> (1)	0	0	0
<i>recén nacido</i> (1)	80	54	0 ³⁵
<i>repartirse o pastel</i> (1)	0	0 ³⁶	0
<i>sen embargo</i> (52)	1952	5683	0
<i>tira e afloxa</i> (1)	6	1	0
<i>xunto con</i> (17)	1165	3336	0
Calcos estruturais de compostos léxicos e sintagmas univertalizados do español			
<i>calqueira</i> (25)	137	58	1
<i>cumpleanos</i> (5)	8	12	2
<i>enganabobos</i> (1)	0	2	0
<i>sobremaneira</i> (2)	18	48	0
Calcos estruturais de palabras derivadas e outras palabras complexas do español			
<i>alonxar</i> (1)	717	463	0
<i>cedo</i> (1) (adx.)	19	17	0
<i>chamativo</i> (8)	78	206	0
<i>chear</i> (7)	14	0	0
<i>conlevar</i> (33)	150	317	0
<i>custe</i> (4)	109	260	0

³⁴ Un dos exemplos do *CORGA* é *comerse ao mundo*. Ademais, hai 15 exemplos de *comer o mundo*, sen clítico reflexivo, no *TILG* e 42 no *CORGA*.

³⁵ Aparece un exemplo de *recén*, co mesmo valor, pero sen o segundo elemento.

³⁶ Rexístranse dous exemplos de *repartir o pastel*.

TÁBOA 39. Continuación

	TILG	CORGA	TLPGP
<i>danino</i> (1)	3	0	0
<i>doencia</i> (1)	81	7	0
<i>fornear</i> (1)	3	17	0
<i>lonxano</i> (4)	1408	701	0
<i>novidoso</i> (11)	72	390	0
<i>rechea</i> (1)	28	17	1
<i>recheo</i> (adx.) (1)	127	123	0
<i>troque</i> (1)	166	24	0

TÁBOA 40. Presenza no *TILG*, *CORGA* e *TLPGP* de formas adaptadas do español

	TILG	CORGA	TLPGP
ELEMENTOS CONSONÁNTICOS			
Padrón <g[e,i]>, <j> / <x>			
<i>alexar</i> (2)	104	8	0
<i>aconsexable</i> (1)	0	0	0
<i>antoxar</i> (1)	210	17	9
<i>antoxo</i> (2)	289	22	9
<i>artiluxio</i> (5)	82	41	0
<i>autoperoxudicar</i> (1)	0	0	0
<i>axeno</i> (2)	13	4	0
<i>desventaxa</i> (2)	36	16	0
<i>enaxenación</i> (1)	7	14	0
<i>engranaxe</i> (2)	30	46	0
<i>exe</i> (2)	2	1	0
<i>imaxen</i> (1)	438	83	0
<i>inxenieiro</i> (3)	212	10	0
<i>inxeniero</i> (1)	16	0	1
<i>marxen</i> (1)	96	26	0
<i>orixen</i> (2)	433	78	1
<i>paradóxicamente</i> (2)	64	85	0
<i>parexa</i> (1)	667	36	26
<i>perxudicar</i> (33)	206	121	0

TÁBOA 40. Continuación

	TILG	CORGA	TLPGP
<i>perxudicial</i> (18)	92	53	0
<i>reflexar</i> (12)	507	514	0
<i>rexir</i> (18)	436	104	2
<i>semexante</i> (1)	40	3	1
<i>sumerxir</i> (1)	70	44	0
<i>surximento</i> (3)	3	4	0
<i>surxir</i> (5)	363	112	0
<i>suxerente</i> (2)	63	143	0
<i>afloxar</i> (1)	53	11	5
<i>trampantoxo</i> (2)	0	0	0
<i>ventaxa</i> (5)	388	126	1
<i>xeneración</i> (2)	500	151	0
<i>xeneral</i> (1)	928	83	0
<i>xeneralmente</i> (1)	106	21	0
<i>xenerar</i> (10)	135	183	0
<i>xoven</i> (26)	821	154	1
<i>xuez</i> (1)	1193	97	8
<i>xuglar</i> (1)	32	0	0
<i>xuzgar</i> (4)	417	62	0
Padrón <n>intervocálico / <ñ>			
<i>entreteneamento</i> (2)	0	1	0
<i>manteñer</i> (1)	5	3	0
<i>teñer</i> (2)	0	1	0
<i>sosteñer</i> (1)	26	0	0
Padrón <n>, <l> intervocálico / Ø			
<i>entrear</i> (1)	1	2	0
<i>soer</i> (25)	1103	516	0
Padrón <ll> / <l>			
<i>desarrollar</i> (1)	12	0	0
<i>elo</i> (10)	1338	224	0
Padrón <ll> / <i>			
<i>estreia</i> (1)	0	0	0

TÁBOA 40. Continuación

	TILG	CORGA	TLPGP
Padrón <pl> / <pr>			
<i>emprego</i> (1)	59	0	0
Padrón <uct> / <ut>			
<i>produtividade</i> (1)	0	0	0
ACENTUACIÓN			
Padrón <ui> / <uí>			
<i>disminuír</i> (2)	85	36	-- ³⁷
<i>huír</i> (2)	0	0	--
DITONGOS			
Padrón <ie> / <e>			
<i>aislamento</i> (1)	16	0	1
<i>calquer</i> (30)	1559	551	2
<i>entendimento</i> (1)	122	24	0
<i>hincapé</i> (1)	0	0	0
<i>mentras</i> (8)	3232	770	16
<i>pimento</i> (2)	138	12	23
<i>razonamento</i> (1)	70	44	0
<i>recén</i> (1)	836	1095	1
<i>reconocimento</i> (1)	27	6	0
<i>rendimento</i> (1)	161	62	0
<i>replanteamento</i> (1)	0	1	0
Padrón <cu> / <ca>			
<i>calquer</i> (30)	1559	551	2
Padrón <ue> / <o>			
<i>troque</i> (1)	166	24	0
Padrón <-d> / <-de>			
<i>antigüidade</i> (2)	104	31	0
<i>cualidade</i> (1)	352	524	0 ³⁸
<i>idade</i> (1)	1159	182	1
<i>esclavitude</i> (5)	7	7	0 ³⁹

³⁷ Dado que o TLPGP recolle textos orais, non se realizan estas dúas buscas onde a adaptación afecta á acentuación gráfica.

³⁸ Recóllese un exemplo de *cualidá*.

³⁹ Recóllese *esclavítu* en dous concellos.

TÁBOA 40. Continuación

	TILG	CORGA	TLPGP
<i>falsedade</i> (1)	60	32	0
<i>mitade</i> (1)	428	165	0 ⁴⁰
<i>mortandade</i> (1)	52	52	0
<i>novedade</i> (4)	307	77	0 ⁴¹
<i>pubertade</i> (2)	41	30	0
<i>rentabilidade</i> (1)	47	48	0
<i>solidaridade</i> (1)	211	144	0
Padrón <-ero,a> / <-eiro,a>			
<i>cocineiro</i> (1)	0	0	0
<i>dineiro</i> (1)	0	0	0
<i>financieiro</i> (4)	112	356	0
<i>ganadeiro</i> (2)	18	2	0
<i>hosteleiro</i> (12)	36	90	0
<i>inxenieiro</i> (3)	212	10	0
<i>perecedeiro</i> (1)	27	16	0
<i>placenteiro</i> (1)	10	13	0
<i>terneira</i> (2)	21	13	2
<i>terneiro</i> (1)	12	2	4
Padrón <-imient> / <-ement>			
<i>desvanecemento</i> (1)	8	8	0
<i>entretenemento</i> (1)	30	8	0
Padrón <-ano, -ana> / <-án, -á>			
<i>cotidiá</i> (2)	3	4	0
<i>cotidián</i> (2)	40	10	0
Padrón <-ear> / <-exar>			
<i>plantexar</i> (16)	302	692	0
Padrón <-ble> / <-bel>			
<i>rentábel</i> (2)	6	26	0
Padrón <-dumbre> / <-dume>			
<i>incertidume</i> (1)	131	132	0

⁴⁰ Rexístranse seis exemplos de *mitá* e un de *mitai*.

⁴¹ *Novedá* recóllese en catro obras.

TÁBOA 40. Continuación

	TILG	CORGA	TLPGP
Padrón <-en> / <-e>			
<i>resume</i> (1)	318	180	0
Padrón <-ería> / <-aría>			
<i>ganadaría</i> (1)	0	0	0
Padrón <-illo> / <-iño>			
<i>gusaniño</i> (1)	1	0	0

TÁBOA 41. Presenza no *TILG*, *CORGA* e *TLPGP* de combinacións de galego e castelán sen padrón regular

	TILG	CORGA	TLPGP
<i>afición</i> (1)	4	7	0
<i>aficionar</i> (1)	0	2	0
<i>conecer</i> (1)	1234	86	1
<i>coñocer</i> (1)	176	12	0
<i>costo</i> (3)	36	35	0
<i>desvantaxa</i> (2)	0	0	0
<i>desventaxe</i> (1)	2	0	0
<i>enxeniero</i> (1)	1	0	0
<i>expremir</i> (1)	0	0	0
<i>inxeniería</i> (1)	2	3	0
<i>inxenheiro</i> (1)	2	0	0
<i>objecto</i> (2)	0	0	0
<i>padar</i> (1)	0	0	0
<i>posúer</i> (1)	2 ⁴²	3 ⁴³	0
<i>sumerxer</i> (1)	2	1	0
<i>vantaxa</i> (2)	2	4	0
<i>ventaxe</i> (3)	48	6	0
<i>xurxir</i> (1)	7	1	0

⁴² Os dous exemplos son de *posuer*.

⁴³ Os tres exemplos son de *posuer*.

TÁBOA 42. Presenza no *TILG*, *CORGA* e *TLPGP* de hipergaleguismos

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
ELEMENTOS CONSONÁNTICOS			
Padrón <n> intervocálico / <ñ>			
<i>destiñar</i> (1)	252	16	0
<i>gaña</i> (7)	123	206	0
<i>mañeira</i> (1)	0	0	0
<i>mediciña</i> (2)	203	41	1
<i>oficiña</i> (1)	121	12	1
<i>pisciña</i> (1)	0	0	0
<i>vaciña</i> (1)	0	0	0
Padrón <n> intervocálico / Ø			
<i>abandoar</i> (5)	405	163	0
<i>almacear</i> (1)	8	0	0
<i>destroar</i> (1)	2	1	0
<i>persoaxe</i> (7)	505	92	0
<i>zoa</i> (1)	10	15	0
Padrón oclusiva + <l> / oclusiva + <r>			
<i>acesibre</i> (1),	0	0	0
<i>compricado</i> (2)	129	6	0
<i>frotar</i> (1)	95	30	0
<i>notablemente</i> (1)	16	0	0
<i>pranta</i> (2)	799	87	0
Padrón <ñ> intervocálico / <n>			
<i>estrano</i> (3)	1252	105	0
DITONGOS			
Padrón <io> / <o>			
<i>comerzo</i> (3)	39	0	0
<i>desperdizo</i> (1)	0	0	0
<i>desprestixo</i> (2)	2	1	0
<i>nazón</i> (1)	301	15	0
<i>prestixo</i> (1)	31	5	0
<i>prestixoso</i> (1)	8	1	0
<i>razonamento</i> (1) (co valor de racionamento)	0	0	0

TÁBOA 42. Continuación

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>versonar</i> (1)	0	0	0
Padrón <ia> / <a>			
<i>aprezar</i> (2)	40	61	0
<i>convivenza</i> (1)	23	0	0
<i>desgrazadamente</i> (1)	6	23	0
<i>influenza</i> (6)	88	37	0
<i>referenza</i> (1)	73	3	0
<i>vivenza</i> (1)	10	5	0
Padrón <ie> intervocálico / <e>			
<i>concenciar</i> (1)	0	0	0
<i>inconvenente</i> (1)	23	5	0
<i>ingredente</i> (5)	16	11	0
TERMINACIÓNS			
Padrón <-ero,a> / <-eiro,a>			
<i>austeiro</i> (1)	5	0	0
<i>calqueira</i> (25)	137	58	1
<i>meiro</i> (1)	0	0	0
<i>sinceiro</i> (2)	229	31	0
<i>sinceiramente</i> (2)	33	8	0
Padrón <-iente>, <-yente> / <-inte>			
<i>influínte</i> (5)	30	97	0
<i>ingredinte</i> (1)	0	2	0
<i>nutrinte</i> (1)	4	11	0
Padrón <-ano, -ana> / <-án, -á>			
<i>humán</i> (2)	5183	240	0
<i>semá</i> (1)	2675	66	2

TÁBOA 43. Presenza no *TILG*, *CORGA* e *TLPGP* de voces creadas por analogía con outras formas galegas

	TILG	CORGA	TLPGP
Formas creadas por analogía con outras voces da familia de palabras ou de familias formalmente próximas			
<i>aprezar</i> (2)	40	61	0
<i>desgrazadamente</i> (1)	6	23	0
<i>expoñencialmente</i> (1)	1	3	0
<i>gaña</i> (7)	123	206	0
<i>gañancia</i> (1)	50	19	0
<i>gañanza</i> (2)	2	0	0
<i>inconvente</i> (1)	0	0	0
<i>inconvinte</i> (8)	6	21	0
<i>obsolencia</i> (1)	0	0	0
<i>persoaxe</i> (6)	505	92	0
<i>xeralizar</i> (3)	108	213	0
Formas creadas por analogía con outras voces do paradigma flexivo			
<i>elo</i> (10)	1338	224	0
<i>entreteñemento</i> (2)	0	1	0
<i>teñer</i> (2)	0	1	0
<i>conseguer</i> (1)	12	9	0
<i>consomer</i> (1)	0	0	0
<i>descobrir</i> (1)	282	48	0
<i>locer</i> (1)	1	0	0
<i>manteñer</i> (1)	5	3	0
<i>seguer</i> (1)	22	14	0
<i>sosteñer</i> (1)	26	0	0
Formas creadas con alomorfos dos sufixos estándar			
<i>influenza</i> (6)	88	37	0
<i>influínte</i> (5)	30	97	0
<i>convivenza</i> (1)	23	0	0
<i>nutrinte</i> (1)	4	11	0
<i>referenza</i> (1)	73	3	0
<i>vivenza</i> (1)	10	5	0

TÁBOA 43. Continuación

	TILG	CORGA	TLPGP
Formas creadas mediante sufixación con competidores coa mesma raíz			
<i>asombrante</i> (1)	1	0	0
<i>bolsátil</i> (1)	0	0	0
<i>diferenciativo</i> (1)	0	0	0
<i>divulgamento</i> (1)	0	0	0
<i>emplataxe</i> (1)	0	0	0
<i>estourido</i> (1)	31	142	0
<i>famosidade</i> (1)	0	0	0
<i>inmemorable</i> (1)	2	1	0
<i>irrecoñecido</i> (1)	0	0	0
<i>lalinés</i> (1)	1	8	0
<i>monotonidade</i> (1)	0	0	0
<i>sacrificación</i> (1)	0	0	0
<i>sacrificante</i> (1)	5	0	0
<i>vanguardoso</i> (1)	0	0	0
Formas creadas mediante derivación sen competidores coa mesma raíz			
<i>acadamento</i> (1)	17	3	0
<i>malgaste</i> (1)	0	3	0
<i>mobilístico</i> (1)	0	0	0
<i>recolleitar</i> (1)	25	11	2
Formas cuxa forma non se corresponde co significado buscado			
<i>autoconsumismo</i> (1)	0	0	0
<i>contrapiar</i> (1)	0	0	0
<i>innaturabilidade</i> (1)	0	0	0
<i>intercomercial</i> (1)	0	0	0
Combinacións «imposibles» de bases e afixos			
<i>moderidade</i> (1)	0	0	0
<i>producidade</i> (1)	0	0	0
Voces cuxa forma se pode explicar pola recorrencia de certas terminacións			
<i>lonxetividade</i> (1)	0	0	0
<i>populariedade</i> (2)	0	0	0
<i>tercialización</i> (1)	0	1	0

TÁBOA 44. Presenza no *TILG*, *CORGA* e *TLPGP* de voces non estándar presentes no galego popular

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>fruto</i> (4)	337	46	11
<i>atrais</i> (3)	93	13	1
<i>bon</i> (3)	4320	641	11
<i>conecer</i> (1)	1234	86	1
<i>coñocer</i> (1)	176	12	0
<i>millor</i> (4)	7216	831	11
<i>tar</i> (3)	2705	48 ⁴⁴	11
<i>terneira</i> (2)	21	13	2
<i>terneiro</i> (1)	12	2	4
<i>fruta</i> (2)	121	29	14
<i>misturar</i> (2)	734	273	1
<i>pimento</i> (2)	138	12	23
<i>buche</i> (1)	1	12	28
<i>caír</i> (1)	3052	430	19
<i>cuadro</i> (1)	321	82	40
<i>cual</i> (1)	1012	95	2
<i>cuanto</i> (1)	459	70	3
<i>tu</i> (1)	2737	310	1
<i>millorar</i> (1)	325	57	6
<i>podrir</i> (1)	9	0	2
<i>recolleitar</i> (1)	25	11	2
<i>semá</i> (1)	2675	66	2

⁴⁴ No *CORGA* só se buscaron as formas do presente, posto que as do pretérito están lematizadas baixo o verbo *ter* e cumpriría levar a cabo a súa localización manual.

TÁBOA 45. Presenza no *TILG*, *CORGA* e *TLPGP* de voces non normativas presentes no galego elaborado

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>soer</i> (25)	1103	516	0
<i>voltar</i> (8)	5030	2954	0
<i>nen</i> (5)	2276	1199	0
<i>blogue</i> (1)	0	37	0
<i>clube</i> (1)	249	492	0
<i>evoluír</i> (1)	35	70	0
<i>nembargantes</i> (1)	590	610	0
<i>respostar</i> (1)	540	631	0

TÁBOA 46. Presenza no *TILG*, *CORGA* e *TLPGP* de estranxeirismos non adaptados

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>tablet</i> (2)	0	163	0
<i>táblet</i> (1)	0	0	0
<i>gasoil</i> (1)	33	25	0
<i>gasioi</i> (1)	0	0	0
<i>amateur</i> (1)	20	80	0
<i>yankee</i> (1)	2	13	0

TÁBOA 47. Presenza no *TILG*, *CORGA* e *TLPGP* de unidades que combinan elementos de varias palabras ou expresións galegas

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>a saber Deus</i> (1)	0	0	0
<i>adedicar</i> (1)	5	0	0
<i>dar lugar a cabo</i> (1)	0	0	0
<i>indixesta</i> (subst.) (1)	0	0	0
<i>non entanto</i> (1)	7	17	0
<i>remontar a vista</i> (1)	0	0	0

TÁBOA 48. Presenza no *TILG*, *CORGA* e *TLPGP* de unidades léxicas non estándar con diferenzas que van máis alá do xénero e da acentuación con respecto ao estándar cuxa orixe non é atribuíble ás razóns sinaladas nas táboas previas

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>estrea</i> (1)	0	0	0
<i>fugaxidade</i> (1)	0	0	0
<i>lumen</i> (1)	2	0	0
<i>reixear</i> (1)	0	2	0

TÁBOA 49. Presenza no *TILG*, *CORGA* e *TLPGP* de palabras con xénero non estándar

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
Palabras con asignación inadecuada de xénero por influencia do castelán⁴⁵			
<i>arte</i> (m) (35)	682	152	3
<i>costume</i> (f) (22)	360	27	0
<i>labor</i> (f) (15)	244	156	1
<i>leite</i> (f) (15)	29	5	7
<i>canle</i> (m) (5)	90	71	1
<i>orixe</i> (m) (4)	264	43	0
<i>viaxe</i> (m) (4)	779	147	3
<i>porcentaxe</i> (m) (3)	19	20	0
<i>análise</i> (m) (2)	48	49	0
<i>engranaxe</i> (m) (2)	0	6	0
<i>equipaxe</i> (m) (2)	24	9	0
<i>marxe</i> (m) (2)	98	75	0
<i>aprendizaxe</i> (m) (1)	26	24	0
<i>arma</i> (m) (1)	5	28	0
<i>color</i> (m) (1)	152	25	4
<i>emplataxe</i> (m) (1)	0	0	0
<i>linguaxe</i> (m) (1)	315	4	0

⁴⁵ Nesta táboa, e para o *CORGA* e o *TILG*, ofrécense unicamente os datos correspondentes aos resultados da busca determinante do xénero non estándar seguido do substantivo correspondente, de tal xeito que non se pode asegurar que as cifras ofrecidas dean conta de todos os casos de empregos con asignación de xénero non estándar. No caso do *TLPGP*, debe terse en conta que non todas as obras ofrecen xénero, expresións ou exemplos que dean conta do xénero da palabra buscada.

TÁBOA 49. Continuación

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
<i>paisaxe</i> (m) (1)	189	6	0
<i>sal</i> (f) (1)	132	27	0
<i>sangue</i> (f) (1)	110	11	2
Palabras con asignación inadecuada de xénero por analogía			
<i>interese</i> (f) (3)	0	3	0
<i>auxe</i> (f) (2)	5	5	0
<i>personaxe / persoaxe</i> (f) (con referente do sexo masculino ou masculino e feminino) (3)	70	Máis de 100 ⁴⁶	0
Palabras cuxa asignación inadecuada de xénero non se explica nin por transferencia nin por analogía			
<i>sabor</i> (f)	3	0	0
<i>Internet</i> (f)	2	20	0

TÁBOA 50. Presenza no *TILG*, *CORGA* e *TLPGP* de palabras con acentuación non estándar

	<i>TILG</i>	<i>CORGA</i>	<i>TLPGP</i>
Palabras con acentuación non estándar por influencia do español			
<i>élite</i> (7)	58	58	0
<i>textil</i> (5)	112	51	0
<i>coltán</i> (2)	0	4	0
<i>periodo</i> (3)	177	174	0
<i>vértixe</i> (1)	6	17	0
Palabras con acentuación non estándar explicable por outros motivos			
<i>Nóbel</i> (1)	2	18	0
<i>nóvel</i> (1)	0	1	0
<i>canón</i> (4)	0	3	0

⁴⁶ Dada a elevada presenza dos lemas *personaxe / persoaxe* no *CORGA* non levamos a cabo a análise detallada de todas as formas e do seu contexto para determinar o sexo do referente. En todo caso, tal e como se indica hai máis de 100 exemplos do substantivo precedido de determinante feminino para facer referencia ben a homes, ben a homes e mulleres conxuntamente.

A posición do pronome persoal átono con formas verbais finitas no corpus *CORTEGAL*

XOSÉ ANTONIO FERNÁNDEZ SALGADO
Universidade de Vigo

1. INTRODUCCIÓN

Un dos trazos máis caracterizadores da sintaxe do galego é, sen dúbida, que o pronome persoal átono pode clitizar situándose adxacente ó verbo tanto despois del (*énclise*) coma antes (*próclise*), como se ve no exemplo *En cambio parécen<me> estupendos eses intercambios que <se> fan entre países para compartir*¹. A posición do clítico no galego estándar non é libre, senón que está gobernada por unha serie complexa de factores, como son (i) os de se tratar dunha forma verbal finita ou non; (ii) de ser o enunciado unha oración independente (*énclise*) ou subordinada (*próclise*); (iii) de se esta é afirmativa (*énclise*) ou negativa (*próclise*); (iv) da presenza na periferia esquerda da oración de constituíntes que poidan ser interpretados informativamente xa como Tópicos ou Temas (*énclise*), xa como Focos (*próclise*), e (v) do acento e entoación que se poidan imprimir e que permitan, por exemplo, unha pausa (*énclise*) ou non (*próclise*) antes da palabra prosódica que conforman o verbo e o clítico (ou o clítico e o verbo).

Esta dupla posibilidade en galego, de que o pronome poida ir enclítico ou proclítico, e tanto con formas verbais finitas coma con infinitivos e xerundios, contrasta coa preferencia doutras linguas románicas pola anteposición pronominal. Por exemplo, castelán, italiano ou catalán prefiren a *próclise*

¹ Todos os exemplos están tirados de *CORTEGAL* respectando a versión final do estudante. No caso de a oración non comezar despois de punto, o exemplo introdúcese con [...].

agás co imperativo e coas formas non finitas. Pola súa parte, o francés ofrece unha próclise máis xeneralizada ó aparecer o clítico tamén anteposto a infinitivos e xerundios, ben que enclítico co imperativo. Máis extremo é aínda o portugués brasileiro falado, xa que presenta unha próclise xeneralizada, pois o pronome átono vai sempre antes do verbo, sexa unha forma verbal conxugada, un infinitivo, un xerundio e mesmo un participio pasado (Martins, 2016, p. 402).

Pola contra, o galego comparte co portugués europeo (Martins, 2013, 2016) e co asturiano (D'Andrés, 1993; ALLA, 2001, pp. 363-369; González i Planas, 2007) a case totalidade de contextos en que se dá a énclise e a próclise. Así, o pronome persoal átono en galego vai posposto ó verbo finito en oracións non dependentes afirmativas (1a), tamén cando hai antes do verbo algún elemento topicalizado, un adverbio (1b), ou tematizado, un complemento directo (CD) (1c), e coas formas verbais en imperativo (1d):

- (1) a. Levantámo<nos> consumindo e deitámo<nos> consumindo tamén
 b. Aquí, os fillos atópan<se> en idades moi complexas
 c. [...] isto vémo<lo> reflexado nos centos e centos de cociñeiros existentes
 d. A min dá<me> comida normal

Pola contra, o clítico colócase antes do verbo en oracións negativas (2a); en subordinadas introducidas por un nexo, sexa unha conxunción (2b), un relativo (2c) ou un interrogativo (2d); e tamén en oracións afirmativas cun constituínte focalizado antes do verbo, que poden ser algúns adverbios (2e), cuantificadores (2f) ou calquera outro constituínte, coma o complemento indirecto (CI) (2g):

- (2) a. Non <se> trata de ser «simples traficantes de comida»
 b. Esto penso que <se> debe ás ganas de innovar e de facer pratos
 c. Hoxe en día, son moitas as persoas que <se> adican á cociña
 d. [...] ata os cociñeiros suben vídeos explicando como <se> fai comidas determinadas
 e. Esto xa <o> fan as propias tendas
 f. [...] todos e todas <nos> vemos afectados por unha inmensa cantidade de publicidade
 g. [...] Polo tanto, a todos nós <nos> interesa que este fenómeno perviva

Co infinitivo e xerundio existe certa liberdade de colocación en galego e, así, o pronome tanto pode ir enclítico, o máis habitual (3a-d), coma proclítico, posición esta menos frecuente (3e-f) (de feito, en *CORTEGAL* non localizamos ningunha próclise con xerundio e só 9 con infinitivo):

- (3) a. [...] basta con preguntar<lle> a cinco nenos
 b. [...] sometidos ó consumismo, a gastar por gastar e a non parar<nos> a pensar
 c. [...] o fin segue a ser o mesmo, pagues máis ou menos [...]: terminar coméndo<o>
 d. [...] facéndo<nos> querer sempre algo máis e non permitíndo<nos> ser felices
 e. [...] ou busca facelo tanto para <se> realizar como para ter un salario
 f. Pero aínda que nos comprendan teñen que <nos> marcar unhas directrices a seguir

A anteposición do clítico co infinitivo e xerundio dáse cando van precedidos dun elemento proclisador, como pode ser unha negación, unha preposición (3e) ou un nexo (3f), neste último caso nunha perífrase. Por iso, en (3a-b) o pronome tamén podería ir proclítico (*basta con <lle> preguntar; non <nos> parar a pensar*) e igualmente co xerundio de (3d) (*e non nos permitindo*). Con respecto ó participio, na variante estándar do galego o pronome átono non clitiza nunca con el.

En todo caso, o pronome átono no galego estándar nunca pode estar en posición inicial absoluta da oración, en cumprimento da lei que Tobler e Mussafia estableceran no século XIX; por exemplo, unha secuencia de próclise como *<Se> *debe ter presente á hora de producir* é agramatical no galego estándar.

Unha visión conxunta da casuística da posición do pronome átono en galego está descrita con precisión na tese de doutoramento de Álvarez Blanco (1980, pp. 402-575) e nas gramáticas galegas máis recentes de Álvarez et al. (1986, pp. 183-206), Freixeiro Mato (2000, pp. 165-185, pp. 694-715) ou Álvarez e Xove (2002, pp. 558-570). Os traballos de Prieto Alonso (1986), Xove (1991) e Freixeiro Mato (2000, pp. 694-715; 2003, pp. 95-102) céntranse especificamente en explicar a posición do clítico en oracións independentes a partir da estrutura informativa da oración e das funcións de Tema e Rema e mais de Tópico e Foco, e as repercusións dos procesos de topicalización, tematización e focalización de constituíntes oracionais na periferia esquerda da oración —a énfase da que xa falara Saco Arce (1868, p. 162), o primeiro gramático en dar unhas regras de colocación para o galego—. Para Xove (1991, p. 34), en último caso, son as funcións informativas de Tópico, Tema ou Foco que adquiren os constituíntes da oración as responsables da posición do pronome persoal átono en oracións independentes de verbo finito; en palabras súas:

depende, si, de aspectos fónicos, como a intensificación e a liña tonal, pero estes á súa vez conforman, xunto coa orde de palabras, tan importante logo nesta problemática, o plano da expresión do significado informativo que o falante distribúe segundo a súa intencionalidade comunicativa. O tópico, o tema e o foco son por tanto os responsables últimos da posición que o pronome ocupa en relación ó verbo en galego [...] de forma que a aparición do pronome anteposto en determinados contextos ou posposto noutros non se debe a ningunha categoría de palabras, a ningunha función sintáctica nin a ningún modo oracional particular: o sistema é sempre o mesmo, varía só a frecuencia con que as expresións poden aparecer focalizadas ou topicalizadas.

Outros estudos sobre a posición do pronome átono centráronse na descrición de corpus particulares de galego. Ogando Vázquez (1980) achegou datos sobre a colocación do clítico no galego antigo, en tanto que Meier (1976) o fixo a partir do galego contemporáneo en transcricións escritas de contos populares. Ademais, Sobriño Pérez (2001) contrastou as colocacións do galego co asturiano. De por parte, unha descrición da posición coas formas non finitas pode verse nos capítulos correspondentes das gramáticas xa citadas e especificamente no traballo de Álvarez (1996), en tanto que para a variación nos complexos verbais pode verse Santos Raña (2018).

Neste capítulo examinamos a colocación do pronome átono con respecto ás formas verbais finitas a partir dunha mostra de arredor de 1600 casos tirados do corpus *CORTEGAL*, que se describe en Álvarez de la Granja e Rodríguez Rodríguez neste libro. O noso obxectivo é ofrecer unha visión global da posición dos clíticos en textos redactados por estudantes galegos de arredor de 18 anos durante as probas de acceso á Universidade. Os datos que obtemos serán contrastados coa descrición da posición canónica que para o galego estándar describen as gramáticas galegas actuais ás que xa aludimos. O máis interesante será identificar os contextos en que a colocación do clítico diverxe da habitual da variedade estándar escrita.

Así e todo, gustaríanos facer antes algunhas observacións xerais. Na explicación dalgunhas colocacións, cando menos de fondo, cumprirá reparar en dous aspectos sobre o corpus que describimos: dunha banda, a situación de comunicación en que foron escritos os textos e, por outra, o perfil lingüístico do estudantado. Canto ó primeiro, debemos ter en conta que os textos analizados foron creados e redactados durante unha proba académica e iso, obviamente, influíu na planificación do propio texto, o tempo dedicado á redacción e mesmo o proceso da posterior revisión. Ademais, trazos lingüísticos que frecuentemente pairan sobre os textos proceden da lingua oral; poderíase dicir que nalgún caso analizamos enunciados escritos oralizados

(ou enunciados orais escritos), o que se reflicte en aspectos como a propia organización informativa, tan importante para explicar a posición dos clíticos, e noutros como, por exemplo, a concordancia ou a puntuación usada.

Por outra parte, algunhas clitizacións só poderán ser explicadas polo perfil bilingüe castelán-galego, activo ou non, dunha boa parte do estudantado e por unha mala aprendizaxe do galego como L2. É importante ter en conta isto, e máis sabendo que as diverxencias de colocación do pronome átono entre galego e castelán poden desencadear interferencias deste naquel e unha adquisición deficiente das regras provocará incertezas e casos de ultracorrección nos usuarios: se a posición non marcada en galego é a *énclise*, ante a dúbida de colocalo antes ou despois, optarán pola posición posverbal. Non coñecemos estudos sobre a adquisición do sistema de colocación pronominal nas nenas e nenos galegos, pero si os hai para o portugués europeo. Costa et al. (2016, pp. 449-450) conclúen que a complexidade do sistema de colocación está na base da súa tardía adquisición por parte das crianzas portuguesas, se se comparan coas francesas, italianas ou españolas, que aprenden linguas de sistemas máis sinxelos; e algo semellante se pode dicir tamén para os aprendentes de portugués europeo como L2, como describe Martins (2016, p. 401). Pola nosa experiencia na docencia de lingua galega nos primeiros cursos universitarios, tamén constatamos que a colocación do pronome átono en galego é un dos aspectos que máis lle custa dominar ó estudantado non galegofalante. E mesmo o galegofalante pode toparse tamén con discrepancias entre a posición do clítico no seu galego oral e a practicada no estándar escrito. Así, nalgunhas subordinadas con nexo causal, como (4), certamente a variante galega estándar esixe a *próclise* (4b); pero no galego oral a *énclise* de (4a) sería unha opción válida se consideramos topicalizada a frase que funciona como CI (subliñada) e a pronunciamos entre pausas cun tonema final descendente. Se fose focalizada (en maiúsculas), a *próclise* sería obrigada (4c):

- (4) a. [...] durará un longo período, porque a moita xente gústa<lle> innovar nas súas receitas (*énclise posible no galego oral*)
 b. [...] durará un longo período, porque a moita xente <lle> gusta innovar nas súas receitas (*próclise preferida no galego escrito*)
 c. [...] durará un longo período, porque A MOITA XENTE <lle> gusta innovar nas súas receitas (*próclise obrigada no galego oral e escrito*)

Queremos advertir con isto que, na análise realizada, demos con casos de colocación pronominal diverxentes coa posición canónica da escrita que resultan difíciles de lexitimar ou non como válidas para o estándar, pois aínda

que poidan ser anotadas como incorrectas para o galego escrito, non o son en rexistros orais. Dificulta o noso labor de enxuízamento que, xaora, descoñecemos a intención, a entoación e a énfase con que o estudantado concibiu e escribiu cada enunciado; nestes casos só nos resta aplicar o noso propio criterio como lectores do texto e xulgar dende a nosa posición de galegofalantes e gramáticos.

No que segue, describimos primeiro as particularidades do corpus analizado (§2), facemos logo unha avaliación xeral da colocación do clítico (§3), para posteriormente describir separadamente os casos que cualificamos como colocacións canónicas, tanto enclíticas (§4) como proclíticas (§5), das que diverxen e poden ser consideradas como agramaticais en galego (§6) e das diverxentes, pero que non o serían, cando menos nalgunhas variedades (§7). O traballo péchase co capítulo de conclusións, nas que sintetizamos os principais resultados obtidos alén de destacar os contextos en que se producen maiores diverxencias de colocación con respecto ó galego estándar escrito —e que, por tanto, precisarán dunha especial atención no proceso de aprendizaxe—, e coa epígrafe dedicada ás referencias bibliográficas manexadas.

2. O CORPUS ANALIZADO

O corpus *CORTEGAL* contén arredor de 6000 formas pronominais átonas, das cales unhas 550 poden ser anotadas como merecentes dalgún comentario con respecto á súa posición. Quere isto dicir que nunha porcentaxe moi alta, arredor do 90% dos casos, a colocación rexistrada en *CORTEGAL* é coincidente coa habitual da lingua estándar e só nun 10% dos casos habería algunha discrepancia: por exemplo, (5a) é un caso de énclice que no galego estándar vai proclítico; pola contra, (5b) é un caso de próclise que debe ser énclice:

- (5) a. O que antes chamába<se> cociñeiro agora é todo un artista
b. Eses cartos <os> sacamos traballando, grazas a o noso salario

Non obstante, co fin de ter unha visión xeral da posición do clítico en *CORTEGAL* e así poder dispoñer, coma tal, de datos globais sobre porcentaxes de énclice/próclise e da incidencia de cada un dos desencadeadores de próclise, decidimos analizar a posición de 1600 pronomes átonos. Como se ve na Táboa 1, nas persoas con menores ocorrencias tivemos en conta o total ou unha cifra moi próxima, pero nas de máis a escolla da mostra fíxose a partir de eleccións arbitrarias en textos non sempre contiguos. Deixamos fóra

os clíticos con formas verbais infinitas, aínda que si incluímos os casos de complexos verbais en que o pronome clítiza coa formal verbal conxugada. A mostra finalmente examinada foi de 1578 exemplos, que representa arredor dun 33% das formas con verbo finito e un 25,5% do total de clíticos de *CORTEGAL*.

Na Táboa 1 móstranse o número de clíticos analizados clasificados pola persoa e a porcentaxe que representa sobre o total de clíticos con formas finitas. Debido á desigualdade de ocorrencias, como vai dito, nalgúns casos analizamos todos ou case todos os exemplos con formas finitas (*me, che, te, vos*), noutros unha porcentaxe considerable, arredor do 60% de exemplos (*o, a, os, as, lle, lles*), e unha menor sobre o total no caso de *nos* e *se*, con preto do 40% e un 16% do total, aínda que son os clíticos que achegan o maior número de casos.

TÁBOA 1. Casos totais de pronomes átonos analizados con respecto ó total de *CORTEGAL*

Persoa	Pronomes	Casos analizados con formas finitas	Casos totais con formas finitas	% analizada de formas finitas	Pronomes totais
P1	me	111	111	100%	118
P2 Ac.	te	55	56	98,2%	86
P2 Dat.	che, cho	62	64	96,8%	68
P3 Ac. sg./pl.	o, a, os, as	359	584	58,4%	1006
P3 Dat. sg./pl.	lle, lles, llo, lla llos, llelos	260	414	62,8%	524
P4	nos, nolas, nolo	285	729	39%	1025
P5	vos	1	1	100%	2
P3/P6	se, se lle, se lles, se llo, se a, se o, se me, se nos	445	2805	15,8%	3338
Total		1578	4764	33,1%	6167

Na mostra analizada, a P3/P6 (*se* seguido ou non doutros pronomes) é a que achega maior número de casos, o 28.2%, logo, a P3 Ac. co 22,7%, a P4 co 18%, a P3 Dat. co 16,5%, a P1 co 7%, a P2 Dat. co 3,9%, a P2 Ac. co 3,4% e, finalmente, a P5, que só achega un exemplo.

3. A posición do pronome átono en *CORTEGAL*

As distintas gramáticas do galego (Álvarez et al., 1986, p. 183; Freixeiro Mato, 2000, p. 166; Álvarez e Xove, 2002, p. 558) coinciden en afirmar que a posición non marcada do galego é a énclise ó verbo e só en determinados contextos o clítico se despraza á posición preverbal, como poden ser as predicacións negativas, as oracións subordinadas ou o cambio na estrutura informativa.

Na Táboa 2 poden verse os recontos globais dunha e outra posición na nosa mostra:

TÁBOA 2. Casos totais de énclise, de próclise e de énclise e próclise á vez con formas verbais finitas

Posición	%	Nº de casos
Énclise	35,1%	555
Próclise	64,8%	1022
Énclise e próclise	0,06%	1
Total	100%	1578

Como se pode ver, a anteposición do pronome ó verbo, cun 64,8%, é o maioritario en *CORTEGAL*, o que sen dúbida se debe ó tipo de textos analizados, nos que dominan as estruturas con subordinación. Como era de esperar, non rexistramos ningún caso de interpolación (isto é, cando en próclise un elemento se coloca entre o clítico e o verbo); pero si anotamos un exemplo (6) en que o pronome átono aparece repetido antes e despois do verbo, cando o esperable sería unha próclise:

- (6) [...] é moi importante hoxe en día xa que se non funcionase ou <o> fixése<o> mal o primeiro mundo habería entrado nunha fonda crise

O caso de (6) se cadra se debe explicar simplemente como anécdota, un erro involuntario por descoido, ben que a forma verbal composta que tamén se recolle no exemplo apunta a unha deficiente competencia en galego. Repetición do clítico con énclise e próclise simultáneas foron atestadas por Costa et al. (2016, p. 441) no seu estudo sobre a adquisición da colocación pronominal en crianzas portuguesas, e algunha vez que outra tamén aparecen no galego oral e mesmo no escrito. Coma tal, na nosa práctica docente universitaria, este curso rexistramos nunha redacción escrita o seguinte caso: *Pero desa volta chamoume a atención a forraxe do libro, pois estaba envolto nun papel de charol, que <me> acheguei<me> a ela.* Así mesmo, a modo de anécdota, pódese comentar que a repetición foi usada estilisticamente polo

escritor X. Cid Cabido no título do relato «Téñome que me irme» da súa obra *Días contados* (1991)²: o efecto expletivo é tamén co clítico de primeira persoa *me* e o xogo de posicións afecta ás tres posibilidades de colocación coas formas verbais perifrásticas cun nexos.

O predominio do número de casos proclíticos en *CORTEGAL* contrasta co dominio da énclice nos datos obtidos por Meier (1976, p. 46) na súa análise de *Contos populares da provincia de Lugo* (1972), como se ve na Táboa 3:

TÁBOA 3. Datos de énclice e próclise no corpus *CORTEGAL* en comparación cos *Contos populares*

Posición	<i>Contos populares</i> (1972)	<i>CORTEGAL</i>
Énclice	60%	35,1%
Próclise	40%	64,8%
Total	2687	1578

Os datos de énclice e próclise invértense nos dous corpus, o que se explica pola distinta tipoloxía textual en cada un deles: textos escritos, supostamente formais, na variedade estándar e de tipo expositivo-argumentativo, no caso de *CORTEGAL*; e textos narrativos transcritos da oralidade popular, nos de Meier. Isto implica predominio duns tipos de oracións sobre outros: máis complexas en *CORTEGAL* e simples nos *Contos*, e máis presenza de subordinación naquel e de coordinación e pausas nestes.

Por outra parte, as porcentaxes de clíticos pospostos e antepostos varían segundo a persoa pronominal (P). Conforme se aprecia na Táboa 4, só na P1 (*me*) os casos enclíticos son máis cós proclíticos, nunha relación de 6/4; na P3/P6 referidas a *se* (seguido ou non doutras formas), as porcentaxes son parellas, aínda que favorables á próclise por pouco; mais, nas restantes persoas, a próclise é maioritaria nunha relación media aproximada de 7 casos antepostos por só 3 pospostos. O caso da P5 (*vos*) non conta, pois só hai un exemplo.

² Debo esta información a Manuel Puga Moruxa, que traballou coa tradución ó inglés do relato.

TÁBOA 4. Énclise e próclise segundo a persoa pronominal

Persoa pronominal	Pronomes	% Énclise	% Próclise
P1	me	63,4%	36,6%
P2 Ac.	te	32,7%	67,3%
P2 Dat.	che, cho	35,4%	64,6%
P3/Ac. sg./pl.	o, a, os, as	20,5%	79,5%
P3/Dat sg./pl.	lle, lles, llo, llos, llelos	28%	72%
P4	nos, nolas, nolo	37,2%	62,8
P5	vos	100%	0%
P3/P6	se, se lle, se lles, se a, se o se llo, se me, se nos	45,8%	54,2%

Así e todo, cómpre dicir que non todas as colocacións responden ó canon do galego estándar. Na Táboa 5 recollemos porcentaxes sobre a gramaticalidade das posicións pronominais examinadas:

TÁBOA 5. Grao de coincidencia da posición do clítico en *CORTEGAL* con respecto ó galego estándar escrito

Casos de colocación canónica coincidentes coa do galego estándar escrito		90,1% ³ (1422 casos)
	Casos de énclise	31%
	Casos de próclise	69%
Casos de énclises e próclises non canónicas		5,2% (82 casos)
	Casos de énclise por próclise	58,5%
	Casos de próclise por énclise	40,3%
	Caso de énclise e próclise asemade	1,2%
Outros casos diverxentes da posición canónica		4,7% (74 casos)
	Casos de énclise por próclise en subordinadas	87,8%
	Casos de próclise en subordinadas tras pausa	12,2%

Sobre o total de 1578 exemplos analizados, nun 90,1% constatamos unha colocación canónica do pronome átono (31% enclíticos e 69% proclíticos); isto é, esas posicións responden ás regras usadas no galego estándar escrito e

³ Neste cómputo van incluídos tamén os casos de coordinación de subordinadas, que se analizan en §5.2.

poderíamos dicir que son válidas dentro da gramática do galego. Poden servir de exemplo de colocación canónica enclítica os exemplos de (7a-b) e de proclítica os de (7c-d):

- (7) a. [...] o consumo non gobernan as nosas vidas: gobernámo<las> nós
 b. [...] non me chama moito a atención, pois basicamente céntra<se> no análise e estudo
 c. [...] pero case nunca <nos> chega e, en moitos casos, queremos comer máis
 d. [...] a cultura galega actual non está sendo considerada como <lle> corresponde

Así e todo, a Táboa 5 deixa ver que rexistramos un 5,2% de colocacións que poderíamos cualificar de non válidas dentro da gramática do galego: porcentualmente uns poucos casos máis de énclice por próclise (8a) ca de próclise por énclice (8b):

- (8) a. [...] este factor é determinante, xa que ninguén plantéxa<se> o feo que pode ser algo (gal. est. *xa que ninguén se plantexa...*)
 b. Estas dúas accións <as> leva facendo o ser humano case dende a súa aparición (gal. est. *Estas dúas accións lévaas facendo...*)

Ademais, outro 4,7% de colocacións catalogámolas como diverxentes coa posición canónica, pero non agramaticais. En (9a) a posición posverbal seguramente non sexa a preferida ou recomendada na lingua estándar escrita, pois contraviría a regra xeral de que a subordinación esixe próclise. Non obstante, con algunhas causais (9a), a énclice é na oralidade unha das posibilidades se o constituínte que vai entre o nexo *porque* e o verbo o interpretamos informativamente como Tópico ou Tema. Pola contra, no exemplo (9b), se facemos as pausas do inciso, o pronome clitizaría preferiblemente posposto ó verbo:

- (9) a. [...] porque ao famoso de turno gustába<lle> comer en restaurantes carísimos
 b. [...] por tanto unha clase social maior, agás excepcións que, como ben din o autor no texto, <as> hai para calquera regra

4. CASOS DE ÉNCILISE CANÓNICA EN *CORTEGAL*

Como vai dito, os casos de énclice canónica supoñen o 31% da mostra. Esta posición é a habitual en galego en oracións non dependentes afirmativas nas que non haxa procesos de focalización nalgúns dos constituíntes da periferia esquerda oracional. Así o testemuñan tamén os gramáticos galegos. Saco Arce (1868, p. 161) afirma que os pronomes átonos «ordinariamente se posponen al verbo». Álvarez et al. (1986, p. 183) tamén coinciden en que nas

oracións simples e na principal das complexas o xeral é a énclice, pois é a que se «adopta cando non se dá unha serie de circunstancias que provocan o seu dislocamento para a posición anteverbal». Pola súa parte, Freixeiro Mato (2000, p. 166) fala da énclice como «posición non marcada», «a colocación non condicionada ou espontánea» en galego, en tanto que Álvarez e Xove (2002, p. 558) a cualifican como a «posición normal» nas oracións non dependentes.

4.1. En oracións non dependentes afirmativas

Nos casos seguintes temos exemplos de pronomes enclíticos en oracións simples afirmativas (10a-c); cando a forma verbal é un imperativo (10d); tamén nos membros independentes das coordinadas copulativas de *e* (11a), adverbias de *pero* (11b) e disxuntivas de *ou* (11c) —ben que nalgúñas disxuntivas pode darse a próclise (véxase en 5.1)—; e tamén nas xustapostas (11d):

- (10) a. A maioría mándan<os> a paixes subdesenvoltos
- b. Está<selles> a ensinar como deben ser fisicamente
- c. Este texto provocou<me> unhas ideas de acordo coa autora
- d. [...] «quére<te> a ti mesmo»
- (11) a. Levantámos<nos> consumindo e deitamos<nos> consumindo tamén
- b. [...] vía fai pouco estes programas, pero debido á súa monotonicidade, cansáron<me> e aburríron<me>
- c. [...] Este afán purista leva á desaparición dos termos intermedios, ou gusta<che> ou non che gusta
- d. [...] Chama a atención os concursos gastronómicos, hai<nos> a moréas hoxe en día

É tamén habitual a énclice nas oracións principais das tradicionalmente consideradas subordinadas substantivas (12a-b), adxectivas (12c) e adverbias, sexan estas temporais (13a), causais (13b), condicionais (13c), concesivas (13d), consecutivas (13e), comparativas (13f) etc.:

- (12) a. [...] e respónden<che> que queren ser futbolistas ou modelos
- b. [...] que unhas marcas definan a túa calidade de vida parece<me> excesivo
- c. Realmente debería<se> cobrar en función á responsabilidade á que se esté exposto no traballo
- (13) a. [...] case sempre que celebramos algo celebrámo<lo> cunha comida
- b. [...] despois nosos pais compran<nos> os seus xoguetes porque son os que queremos
- c. Se alguén nos ofrece o seu ombreiro para chorar, rexeitámo<lo> polo reloxo carísimo
- d. [...] aínda que eu sexa un deles dígo<lles>, como pretendedes mellorar na vida

- e. ás veces interé^a<lles> tanto que absorben os coñecementos
- f. [...] Dá^mos<lle> máis importancia a ter un Iphone 7 ou unhas Adidas que a estar san

En todo caso, cómpre salientar que en galego o clítico non pode ir en posición inicial absoluta. Ó comezo da oración pode aparecer unha forma verbal finita (14a-d) e antes desta constituíntes topicalizados ou tematizados, habitualmente o suxeito (SUX) (15a-d), pero tamén con frecuencia CD (16a-d) e CI (16e-h) redobrados:

- (14) a. Paré^e<me> tamén moi importante que os galegos sepamos valorar a nosa propia comida
- b. Considé^ro<o> unha forma de tirar os cartos e un desprezo
- c. Dé^be<se> continuar con esta nova tendencia que fai desfrutar a traballadores
- d. Ensíⁿan<nos> a producir e a consumir, pero non a reflectir
- (15) a. Este tipo de sociedade obrí^ga<nos> a ser moito máis competitivos na produción
- b. O que comezou sendo como unha moda pasaxeira, acabou<se> convertendo nun dos puntos máis destacados
- c. Porén outra gran parte rexéⁱa<os> por causa do descoñecemento sobre as patoloxías
- d. [...] e o pai neste caso, está<o> a pagar con a súa filla
- (16) a. Un exemplo moi claro vé^mo<lo> nos restaurantes
- b. Un exemplo desta realidade, podé^mo<lo> ver no texto
- c. Outra liña en auxe na actualidade, seguen <a> aqueles cocíⁿeiros
- d. [...] isto vé^mo<lo> reflexado nos centos e centos de cocíⁿeiros existentes
- e. E iso á xente gú^sta<lle>, gusta<lle> ver un prato ben presentado
- f. [...] a nós, os galegos, coñece<senos> polo cocido
- g. A min persoalmente gó^sta<me> moito probar de cando en vez comida estranxeira
- h. [...] a moita xente, tanto maiores como pequenos, interé^san<lles> a comida típica

Aínda na periferia oracional poden aparecer antes do verbo outros Tópicos temporais e locativos (17a-f), marcadores discursivos e modificadores oracionais (18a-i) e outros incisos, precedidos doutros constituíntes (19a-c):

- (17) a. A día de hoxe bombardé^a<senos> con noticias vanais constantemente
- b. Dende nenos mé^te<senos> na cabeza que se non somos inxenieros
- c. Cada día que pasa, converté^mos<nos> máis e máis nunha sociedade consumista
- d. No pasado, plantá^ban<se> patacas, cebolas e todo tipo de legumbres
- e. Hoxe en día, coa evolución da cociña e o descubrimento de novas especias, pó^de<se> facer unha variedade de pratos
- f. [...] no texto, por exemplo, preséⁿta<nos> a historia de Carmela de Chousa

- (18) a. Aínda así, parece<me> que o autor acerta
 b. Ademais, dá<me> a sensación de que é un mundo interesante
 c. Deste xeito, vai<se> xerando un bucle que eleva ese país gastronOMICAMENTE
 d. [...] porén, sómo<lo> todos dende o berce ata o burato final
 e. Agora ben, eu identifíco<me> coa situación xa que eu tamén fun así
 f. Por último gustaría<me> plantexar unha pregunta
 g. En conclusión, dámos<lle> máis importancia a as apariencias
 h. En definitiva, parece<me> boa idea dar lle unha gran importancia á cocina
 i. A modo de síntese, gustaría<me> remarcar que o consumismo non é tan malo
- (19) a. A cociña, igual que outras artes, bása<se> na imaxinación e nos utensilios
 b. A comida, como calquer outro elemento, encóntra<se> no limbo do tempo
 c. Esta, en moitos casos, aprovéita<se> das modas e das tendencias

O redobro do clítico co CD e CI, en (16), e a énclice son índices de eses constituíntes da periferia esquerda estaren tematizados ou topicalizados. De feito, estes foron os factores que nos levaron a analizar como Temas ou Tópicos e, xa que logo, darmos como canónicas as énclices en que aparecen os demostrativos *isto*, *iso* ou o pronome *elo* en (20) (énclices maioritarias neste caso en *CORTEGAL*). Nalgúns casos, cabería a súa interpretación como Foco e, daquela, o pronome debería ir proclítico, como exemplificamos en (20a) co Foco en maiúscula —e podería ser extensible a algúns outros exemplos desta serie como (20j-m)—:

- (20) a. Isto léva<nos> a falar de consumismo (= *ISTO nos leva a falar de consumismo*)
 b. [...] isto podémo<lo> ver claramente na TV
 c. Isto buscan<o> a maioría das empresas.
 d. [...] e iso sáben<o> ben os que están ao outro lado
 e. Esto débe<se> a que vivimos nunha sociedade capitalista
 f. [...] iso débe<se> a un cambio cultural na sociedade
 g. A sociedade actual é moi consumista, e isto agrandou<se> debido a Internet.
 h. [...] e iso gústa<me> xa que de certo modo estamos reivindicando
 i. [...] e, obviamente, iso beneficia<nos> a nós
 j. [...] e para iso necesíta<se> unha produción deses materiais
 k. Por iso en moitos países quére<se> adoptar a forma que temos
 l. Por iso débe<se> intentar favorecer o conxunto da poboación
 m. Por elo débe<se> apoiar (mediante créditos, asociacións e agrupacións...)

Ademais destes demostrativos, resulta problemático tamén xulgar a colocación do clítico cando antes do verbo aparecen cuantificadores adverbiais, pronominais ou determinantes como *cada* (6 casos), *moita* (1 c.), *ambos* (2 c.), *máis* (2 c.), *todo* (2 c.), *calquera* (1 c.); o adverbio de foco *incluso* (1 c.), dous en *-mente* (*seguramente*, *simplemente*) e outras expresións como *por esta razón* (1 c.). É un grupo que alén da interpretación informativa como

Tema ou Tópico (a elixida por nós e cremos que tamén polos autores e autoras dos textos), para algúns lectores tamén soportarían a focalización, cunha entoación enfática e, por tanto, a próclise do pronome átono. Algúns destes 17 exemplos listámoslos en (21a-j) xunto á posibilidade de focalización (en maiúsculas, a focalización):

- (21) a. Ademáis, cada vez potencia<se> máis a función de produción agraria ou gandeira (= *CADA VEZ se potencia...*)
 b. Cada día máis e máis xente interésa<se> pola cociña, e este interese e increíblemente útil (= *Cada día MÁIS E MÁIS xente se interesa...*)
 c. A partir de aquí moita xente preguntará<se>: E o seu éxito a que se debe ? (= *MOITA XENTE se preguntará...*)
 d. [...] e a todo o mundo encánta<lle> comer os seus platos preferidos (= *TODO O MUNDO lle encanta...*)
 e. Ambos aféctan<nos>, un en canto ao que traballamos e outro en canto ao que adquirimos (= *AMBOS nos afectan...*)
 f. A calquera persona gustaría<lle> comer nalgún dos restaurantes máis prestixiosos (= *A CALQUERA PERSOA lle gustaría...*)
 g. [...] despois pode poñer en práctica a elaboración de ditos platos na súa casa, e incluso pode<llo> transmitir a outra xente (= *INCLUSO llo pode...*)
 h. Esta situación que estase a dar no texto seguramente pasou <lle> a calquer nai o pai (= *SEGURAMENTE lle pasou...*)
 i. Excedentes de materiais que, dado que non os usamos, simplemente tirámo<los> ao lixo (= *SIMPLEMENTE os tiramos*)
 j. Por esa razón produciu<se> o crack bolsatil de o 29 (= *POR ESTA RAZÓN se produciu...*)

Nestes casos Saco Arce (1868, p. 162) dá os cuantificadores *calqueira* e *todo* antes do verbo como causantes de próclise. Concordan Álvarez et al. (1986, p. 186) e Álvarez e Xove (2002, p. 562) con que tamén son proclisadores *calquera* e *todo* e mais *ambos* e *máis*, aínda que logo na exemplificación de *ambos* e *todo* Álvarez e Xove tamén recollen casos de énclice, e para *moito* ofrecen as dúas posibilidades. Para *cada*, Álvarez et al. (1986, pp. 187-188) sinalan que nunca provocan anteposición. Pola súa parte, Freixeiro Mato (2000, p. 169) tamén afirma dos «pronomes» *ambos*, *calquer*, *todos* e dos «adverbios» *moito*, *máis* que cando van antes do verbo non seguidos de pausa provocan a posición preverbal, aínda matizando que ofrecen un comportamento complexo. Xa no portugués europeo, Martins (2016, p. 407) dá *ambos*, *mais*, *todo/tudo* como proclisadores categóricos (ou sexa que exclúen a énclice) e *cada*, *qualquer*, *muitos* coma non categóricos.

4.2. Énclise en oracións subordinadas

As gramáticas galegas reservan algúns parágrafos para comentar a excepción de que nalgunhas subordinadas con nexos o pronome poida ir enclítico. Álvarez et al. (1986, pp. 194-199), Freixeiro Mato (2000, p. 167) e Álvarez e Xove (2002, pp. 565-566) comentan a posibilidade de énclise con algunhas causais, consecutivas, tamén completivas e relativas, e destacan como causantes da posposición o comportamento dalgunhas conxuncións e a quebra da continuidade entre o nexo e o clítico pola distancia entre deles ou unha pausa.

Neste apartado limitarémonos soamente a constatar os casos de subordinadas como (22) en que a énclise é o habitual, como sucede coas causais de *pois* (22a-b) e a consecutiva con *así que* (22c), da que só temos un exemplo⁴:

- (22) a. [...] mercamos aquel que é máis caro pois dá <nos> a sensación de que é de maior calidade
 b. [...] se pode comprar uns zapatos de marca, pois compran <se> uns baratos
 c. Así que hoxe toca <nos> vender os nosos pobos con a gastronomía

Outros casos de énclises en subordinadas rexistrados en *CORTEGAL*, como na completiva *O peor é que os grandes produtores sáben<o>*; na causal *xa que o prezo dos servizos e produtos é<lles> fácilmente asumible*; ou na relativa *unha comida que aos estranxeiros cháma<lles> a atención* comentarémolas en §6.1.3 e §7.1.

5. CASOS DE PRÓCLISE CANÓNICA EN *CORTEGAL*

A próclise é maioritaria en *CORTEGAL*. Dentro do total dos datos analizados na Táboa 2 constitúe o 64,8% e dentro dos casos de colocación canónica, recollidos na Táboa 5, constitúe o 69%.

Certamente, en galego o clítico pasa a posición preverbal sempre e cando na oración haxa un elemento proclisador que provoque ese desprazamento. Entre os gramáticos galegos, Saco Arce (1868, pp. 161-162) adiantouse a sinalar que a posición anteverbal do pronome se produce (1º) cando antes hai algún «pronombre negativo, interrogativo ó conjuntivo» e certos «adjetivos» (*calqueira, pouco, todo*); (2º) cando o verbo vai precedido dalgunha conxunción (exceptuando algunhas como as coordinativas) ou dun adverbio (quitando algúns como os de lugar e tempo); e (3º) cando se antepoñen ó verbo algúns elementos para dar «mayor énfasis a la expresión».

⁴ Pola contra, as dúas consecutivas con *de maneira que* e *ata tal punto que* esixen próclise en (34a-b). En canto ás 6 clasificadas como causais consecutivas con *polo que*, en 4 hai énclise e en 2 próclise, como se pode ver en (65) e (35), respectivamente.

Pola súa parte, Álvarez et al. (1986, pp. 183-184) resumen que a anteposición se dá cando hai presenza «no inicio do enunciado dunha serie de elementos tónicos que non permiten pausa entre si e o verbo» e sinalan que nas «cláusulas subordinadas é de regra a proclise, salvo en certos casos de excepción»; a seguir (1986, pp. 184-189) pormenorizan unha serie de casos tendo en conta factores como a «entoación non enunciativa», os «enunciados negativos», os «precedidos dun indefinido», os «precedidos dun numeral», os «precedidos dun adverbio» (onde distinguen «sempre enclise», «sempre proclise» e «enclise e proclise»).

Tamén Freixeiro Mato (2000, pp. 167-170) salienta contextos semellantes para a anteposición pronominal: (i) en «cláusulas tradicionalmente chamadas subordinadas» introducidas por nexos conxuntivos; (ii) en «cláusulas desiderativas» cun SUX precedendo o verbo; (iii) en «cláusulas enfáticas» en que se quere salientar por focalización o SUX ou calquera outra palabra ; e (iv) cando antes do verbo hai unha serie de palabras non seguidas de pausa, como negacións, «exclamativos e interrogativos», «pronomes identificadores e cuantificadores», e adverbios como os de dúbida, cantidade, «limitación», «inclusión», «identificación» e «presentativos de lugar».

E igualmente, Álvarez e Xove (2002, p. 559) xeneralizan salientando que a próclise é o común nas oracións subordinadas en tanto que nas non subordinadas o clítico sitúase proclítico en «estruturas marcadas por focalización», «oracións con modalidade desiderativa», «oracións interrogativas directas» iniciadas por un pronome interrogativo, «oracións exclamativas iniciadas por unha palabra ou un sintagma exclamativo», «oracións con marca de negación ou afirmación», e oracións que presentan na posición anteverbal «certos adverbios e certos pronomes cuantificadores e identificadores».

Finalmente, volvemos lembrar que tanto Xove (1991) coma Freixeiro Mato (2003, pp. 95-102) salientan que para explicar a próclise en oracións non dependentes debe atenderse sobre todo ós procesos de focalización de constituíntes situados antes do verbo.

5.1. Elementos indutores de próclise

No que compete á nosa mostra de análise, dentro dos casos que cualificamos de colocación canónica, a posición anteverbal do pronome átono prodúcese pola acción dunha serie de proclisadores ou indutores de próclise, que clasificamos na Táboa 6, así como a súa porcentaxe de influencia.

TÁBOA 6. Indutores de próclise canónica en *CORTEGAL*

Proclisadores		Casos de próclise	% do total
Nexos		628	65,6%
	Relativos	314	32,8%
	Conxuncións e locucións conxuntivas	303	31,7%
	Interrogativos	11	1,1%
De negación		222	23,2%
Adverbios focalizados		85	8,9%
Cuantificadores focalizados		20	2,1%
Outros constituíntes focalizados		1	0,1%

Como se aprecia, son os relativos, conxuncións e interrogativos e exclamativos os maiores indutores de próclise no corpus analizado, con case 2/3 dos casos, seguido das palabras que indican negación, case 1/4 dos exemplos, e en moi menor importancia os adverbios e os cuantificadores focalizados preverbaís, que supoñen 1/10 do total. Anecdóticos son os datos doutros constituíntes focalizados que provocan próclise, só un caso.

Cómpre aclararmos que en ocasións pode haber varios posibles proclisadores; nese caso a fichaxe fíxose a favor do proclisador máis próximo do clítico. Así, en (23a) hai ata tres posibles proclisadores, o relativo *onde*, o adverbio *xa* e a negación *non*; neste caso, a próclise adxudicámoslla ó adverbio negativo; en (23b), onde hai a conxunción causal *porque* e o adverbio aspectual *sempre*, adxudicámoslla ó adverbio:

- (23) a. [...] motivalo pra que saque forzas de onde xa non <lle> quedan pra facer algo da sua vida
 b. [...] que din que hai que alcanzar, porque sempre <se lle> bota a culpa ao traballador

Nas epígrafes seguintes, esmiuzamos máis estes datos con cada grupo dos proclisadores.

5.1.1. *Nexos proclisadores*

Nos nexos proclisadores recollemos os relativos, os conxuncionais e os pronomes interrogativos, en total 628 casos. Na Táboa 7, pode verse a lista deles e o tipo coas porcentaxes de ocorrencia:

TÁBOA 7. Lista de proclisadores nexuais en *CORTEGAL*

Nexos	Tipo	% de casos
Que	relativo	36,6%
Que	completivo	17%
O/a/os/as que	relativo	11,4%
Se / si	condicional	6,5%
Cando	temporal	4%
Porque	causal	3,6%
Para que	final	2,7%
Xa que	causal	2%
Como	modal	2%
Tan/tanto(s)/a(s)... como	comparativo	1,1%
Tal e como	modal	1,1%
Como	comparativo	1,1%
Quen	relativo	0,9%
Como	interrogativo	0,9%
Aínda que	concesivo	0,9%
Que	interrogativo	0,6%
Dende que	temporal	0,6%
Senón que	adversativo	0,5%
Onde	relativo	0,5%
Mentras (que)	temporal	0,5%
Coma/como se	comparativo	0,5%
Cal	relativo	0,5%
Unha vez (que)	temporal	0,3%
Sempre que	condicional	0,3%
Polo que	causal consecutivo	0,3%
Máis que	comparativo	0,3%
Como	causal	0,3%
Ou	disxuntivo	0,3%
Tan rápido como	temporal	0,1%
Sen que	modal	0,1%
Sempre e cando	condicional	0,1%
Se ben	concesivo	0,1%

TÁBOA 7. Continuación

Se	completivo	0.1%
Que	imperativo	0.1%
Posto que	causal	0.1%
Por que	interrogativo	0.1%
Por máis que	concesivo	0.1%
Debido a que	causal	0.1%
De maneira que	consecutivo	0.1%
De maneira que	modal	0.1%
Dado que	causal	0.1%
Ata tal punto que	consecutivo	0.1%
Ata que	temporal	0.1%

Os proclisadores máis frecuentes son, destacadamente, os *que* relativos, seguidos dos conxuncionais *que* completivo, *se* condicional, *cando* temporal, os causais *porque* e *xa que*, o final *para que* e o modal *como*, os cales en conxunto supoñen o 85,8% do total. Na Táboa 8 consignamos as estruturas sintácticas en que se deu próclise do pronome átono:

TÁBOA 8. Estruturas sintácticas en que se deu próclise nexual

Estrutura sintáctica	%
Relativa	50%
Completiva	17,4%
Condiciona	7%
Causal	6,4%
Temporal	5,7%
Modal	3,34%
Comparativa	3%
Final	2,7%
Interrogativa	1,7%
Concesiva	1,3%
Adversativa	0,5%
Consecutiva	0,3%
Causal consecutiva	0,3%
Disxuntiva	0,3%

Como era de esperar, son maioritariamente estruturas das tradicionalmente consideradas subordinadas, pero tamén vemos que se dá próclise nos 8 casos das interrogativas directas (24a-b), amais das 3 indirectas (24c):

- (24) a. E o seu éxito a que <se> debe?
 b. Se nos non coidamos do noso, quen <o> fará se non?
 c. Ás veces é irónico como <nos> chama a atención algo cotidiano

E entre as coordinadas, hai próclise tamén en 3 adversativas con *senón (que)* (25a-c) e en 2 disxuntivas exclusivas (25d-e).

- (25) a. [...] non só do seu propio país, se non que <os> proban tamén cando viaxan ao estranxeiro
 b. [...] a citada globalización (grazas a Internet) se non que <nos> atopamos nun momento no que nunca antes se coñecera
 c. [...] polo que non será algo pedecedeiro, senón que <se> converterá en cultura
 d. [...] todo o mundo ten que cociñar ou <se> aprende ou non queda outra
 e. Hoxe en día ou <te> adaptas ás demandas que pide o mercado e intentas expandir o teu negocio ou a empresa morre

Con respecto á próclise das disxuntivas con *ou* (25d-e), Álvarez et al. (1986, p. 194) recollen esta posibilidade de anteposición do pronome «se un elemento exclúe a outro», e igualmente Álvarez e Xove (2002, p. 564). Para o portugués europeo recollen tamén esta posibilidade Brito et al. (2003, p. 856), e tamén a comenta D'Andrés (1993, p. 80) para o asturiano, aínda que cualifica a colocación como «anómala». Na nosa mostra analizada, outro caso de disxuntiva exclusiva presenta pola contra énclice, como xa vimos en (11c).

Por outra parte, entre as tradicionalmente consideradas subordinadas, é interesante separar os casos en que que o nexo vai seguido do clítico (NEXO + CLÍTICO + VERBO) daqueles outros en que entre eles aparecen outros constituíntes oracionais (NEXO + CONSTITUÍNTE + CLÍTICO + VERBO), posto que, como veremos logo en §7.1, neste caso o pronome poderá situarse en posición enclítica ó verbo. Nas estruturas comparativas (26a-c) e modais (26d-f) localizadas, nexos e clíticos van seguidos e poderíamos dicir que, en condicións normais, forman parte da mesma frase entoativa:

- (26) a. [...] facer cousas novas, como <o> foi no seu momento cociñar postres con nitróxeno líquido
 b. [...] pois podemos converternos en escravos do consumo, mais do que <o> somos hoxe
 c. [...] sen que haxa a necesidade de mercar tanto como <o> estamos a facer agora
 d. [...] que eles queiran e tratan a súa familia como <lles> veña en gana

- e. [...] sen modificacións de ningún tipo, tal e como <a> coñecemos dende décadas atrás
 f. [...] e actuar de maneira que <a> melloremos

Pola contra, noutras subordinadas temos exemplos das dúas posibilidades, como recollemos na Táboa 9, que ordenamos segundo o maior número de casos que achega cada tipo:

TÁBOA 9. Secuencias de nexo e clítico en próclise en subordinadas

Tipo de subordinada	NEXO+CLÍTICO+VERBO	NEXO+CONST.+CLÍTICO+VERBO
Relativa	91,1%	8,9%
Completiva	65,1%	34,9%
Condicional	52,3%	47,7%
Causal	62,5%	37,5%
Temporal	75%	25%
Final	30%	70%
Concesiva	87,5%	12,5%
Consecutivas	50%	50%
Causal consecutiva	50%	50%

As porcentaxes da táboa indican que a posibilidade de que haxa un constituínte entre o nexo e o clítico non é igual en todos os tipos de construcións: é maioritaria nas finais, en 7 de cada 10; pero noutras como as relativas e concesivas é minoritaria, só arredor de 1 caso de cada 10; igualmente nas temporais, con 1 caso de cada 4. Nas condicionais son dúas posibilidades parellas e nas completivas e causais, en máis de 6 exemplos de cada 10, o nexo e o clítico van seguidos. Nas consecutivas e causais consecutivas só rexistramos dous casos de cada tipo.

Nos exemplos que veñen a seguir, os (a-b) responden á secuencia NEXO + CLÍTICO + VERBO, e os (c-d) a NEXO + CONSTITUÍNTE + CLÍTICO + VERBO. En (27) son relativas, en (28) completivas, en (29) condicionais, en (30) causais, en (31) temporais, en (32) finais, en (33) concesivas, (34) consecutivas e en (35) as que denominamos causais consecutivas de *polo que*:

- (27) a. [...] non se conforman con calquera cousa que <se lles> poña diante
 b. [...] vivimos nun mundo no cal <se> basa na produción e no consumo
 c. [...] as súas mercancías, o que ás veces <lles> outorga gran peso a todos os niveis
 d. [...] e experimentar, cousa que aos pais <lles> preocupa e crean se moitos conflitos

- (28) a. Creo que <a> chaman a idade do pavo
 b. [...] e isto fai que <o> merquemos cada vez máis e máis
 c. Esta claro que a culpa <a> teñen os fillos
 d. Quizais o peor de todo sexa que esta moda <a> estamos extrapolando
- (29) a. Os pais apoian agora ao seu fillo se <lles> di que de maior quere ser cociñeiro
 b. [...] para a maioría comer é un pracer así que se <o> podes facer é ser o máis «guay»
 c. [...] e moi importante, se nalgun proceso algen <o> fai mal, o resto dos procesos van mal
 d. [...] que son moitas cousas, porque se todo o mundo <o> ten... ha de ser bo
- (30) a. Algunhas persoas desempeñan o oficio de gastrónomo porque <lles> gusta a comida
 b. [...] e a produción manteñen activa a sociedade xa que <a> fai competitiva
 c. [...] excesivamente caros para fardar, ou porque todo o mundo <o> ten
 d. [...] ten tanto éxito porque a toda a xente <lle> gusta comer ben e probar cousas novas
- (31) a. Cando <lles> preguntei por que o mercaran, ningunha me soubo responder
 b. Non parece un plato de comida, mais cando <o> tes diante e <o> probas dáste de conta
 c. Cando algo <nos> gusta, quedámonos con iso durante o maior tempo
 d. [...] mentras que a case ningún outro oficio <se lle> dá dereito a isto
- (32) a. [...] non ten outra fin que para que <a> siga unha consumición
 b. [...] é que a comida sexa de bo ver para que <che> queden ben as fotos nas redes sociais
 c. [...] unhas normas para que no futuro eles <o> fagan cos seus fillos
 d. [...] a nosa sociedade baséase en producir bens para que outro <os> compre
- (33) a. Somos consumidores, e aínda que <o> sabemos, continuamos gastando diñeiro
 b. [...] do consumismo e de produtividade —por máis que <lle> pese a Marx—
 c. Aínda que á sociedade <lle> guste aparentar, a crise obríganos a escoller prioridades
- (34) a. [...] para poder mercar cousas ata tal punto que <as> mercamos sen ter ningunha necesidade
 b. De maneira que isto <nos> deixa cun claro auxe
- (35) a. [...] un fenómeno chamado inflación, polo que <se> produce unha suba masiva dos prezos
 b. [...] seus fillos, polo que, a menudo, <os> fan escoller a que consideran correcta

Na Táboa 10 anotamos os constituíntes funcionais oracionais que se introducen entre o nexa e o clítico (que van subliñados nos exemplos de arriba):

TÁBOA 10. Constituíntes funcionais entre o nexo e o clítico

Tipo de subordinada	SUX	CD	CI	CCIR/ Adxunt.	CCIR-SUX
Completiva	27	3	6	1	
Condicional	14			5	1
Causal	10		3	2	
Final	10			1	1
Relativa	5		4	17	1
Temporal	2	1	1	1	
Consecutivas	1				
Concesiva			1		
Causal consecutiva				1	
Total (%)	69 (58%)	4 (3,3%)	15 (12,6%)	28 (23,5%)	3 (2,5%)

Conforme os datos, estes constituíntes son maioritariamente SUX, como en (29d), CCIRC e outros adxuntos oracionais, coma en (27c) ou (35b), e CI (27d). En moita menor proporción aparecen CD (28c). Poden ser tamén dous os elementos que entremedien, coma en (29c). Informativamente, estes constituíntes están focalizados, o que contribúe tamén á próclise do pronome (alén da presenza do nexo). Algúns exemplos máis incluímolos en (36) e van subliñados:

- (36) a. [...] (cadaquén cociña o que máis <lle> gusta)
 b. [...] e deixaremos de prestarlle atención ás que antes <o> foron
 c. [...] para fomentar certos pratos que para a vida cotiá <nos> resultan chamativos
 d. [...] facendo que cada móbil de 1000 € <lles> coste apenas uns 20 dólares
 e. [...] pois eu creo que iso <lle> dou visibilidade a un sector que non estaba moi recoñecido
 f. [...] penso que a moitas familias <lles> é imposible ser uns «bos consumidores»
 g. [...] porén, como a igualdade <se> produce cando todos temos os mesmos dereitos...
 h. [...] demasiado correcto, dado que moitas persoas <se> adican a elo profesionalmente
 i. [...] podría aumentar a produtividade si á vez a xente <o> consumía
 j. [...] quero dicir con isto que si eu <me> tiro ao río veñen todos detrás
 k. [...] esto seguirá avanzando sempre e cando a capacidade económica <o> permita
 l. [...] prezos de diferentes produtos, sempre que a súa renda <llo> permita
 m. [...] moitos anos máis entre nós ata que unha nova <a> sustitúa

5.1.2. Proclisadores de negación

Son os indutores de próclise no 23,2% dos casos canónicos; os segundos proclisadores máis numerosos tras o macrogrupo dos nexos. Contabilizamos na Táboa 11 as palabras de significado negativo que, en posición preverbal, fan anteceder o pronome, xunta as súas porcentaxes de ocorrencia (222 casos en total). É o contexto de próclise en que se van producir menos anomalías na colocación:

TÁBOA 11. Palabras de negación que provocan próclise

Proclisadores de negación	% de casos
non	85,1%
ninguén	4,9%
nunca	2,7%
nin	1,8%
nadie/naide	1,8%
nada	1,3%
ningún /ningunha	0,9%
nin (tan) sequera	0,9%
xamais	0,5%

O maior número de ocorrencias corresponden a casos do adverbio *non* (37a). Tamén entran aquí outros adverbios como *nunca* (37b) e *xamais* (37c), indefinidos como *ninguén* (37d), *nada* (37e), *ningún/ningunha* (37f), *nadie/naide* (37g), a locución adverbial *nin (tan) sequera* (37h) e a conxunción coordinante *nin* (37i):

- (37) a. [...] queren mercar algo ou ir a algún sitio pero os seus pais non <llo> permiten
 b. [...] ou non máis diñeiro e de iso nunca <nos> temos que olvidar
 c. [...] xa que no resto da historia xamáis <se> viu tal avance en tan pouco tempo
 d. [...] despois de traballar moito a ningúen <lle> apetece cociñar
 e. É certo que se non hai consumidores de nada <lles> serve aos produtores fabricar produtos
 f. Cando lles preguntei por que o mercaran, ningunha <me> soubo responder
 g. [...] ou cando inventaron o lume sen que nadie <lles> dixese nada
 h. [...] verdadeiro problema e que xa nin tan sequera <nos> importa
 i. [...] a atención nun produto que nin <se> consome nin <se> produce, só viaxa

Nun bo número de casos estes proclisadores negativos coinciden con outros indutores de anteposición pronominal, como vemos en (38), aínda que os de negación son os que están máis achegados ó clítico:

- (38) a. Se a gran maioría non <nos> encargamos da produtividade, que esperamos facer?
 b. [...] rexeitan o resto de traballos porque non <lles> atrae
 c. [...] pero mentres que non <o> faga seguirá sendo unha técnica de cociña moi cara
 d. [...] a pesar de que moitos non <o> precisen porque o que teñen está case novo
 e. [...] pra que saque forzas de onde xa non <lle> quedan pra facer algo da sua vida
 f. [...] contra á xente tradicionalista, á que non <lle> gusta, directamente, este tipo de «inventos»
 g. [...] por exemplo os platos con aguacate que antes case ninguén <os> coñecía e agora sí
 h. [...] e quizais non <lles> guste que a súa profesión sexa coñecida solo pasaxeiramente

5.1.3. Adverbios focalizados

Adverbios como *só*, *tamén* ou *sempre*, cando van en posición preverbal, son os causantes da próclise do pronome átono no 8,9% do total de anteposición nos casos canónicos. Trátase de adverbios focalizados, que forman parte da mesma frase entoativa có clítico e que gramaticalizaron a facultade de provocar a anteposición por mor da altísima frecuencia con que isto sucede no galego; isto é, non adoitan ser Tópico. Na Táboa 12 pode verse a lista de adverbios proclisadores na mostra analizada e a porcentaxe que representa con respecto ós 85 casos consignados.

TÁBOA 12. Adverbios focalizados que provocan próclise

Proclisadores adverbiais	Nº de casos
só / soamente	26%
tamén	24%
sempre	11,8%
si / si que	9,8%
xa	9,8%
si que	3,8%
ben	2,4%
así	2,4%
probablemente	2,4%
incluso	2,4%
soamente	2,4%
aínda	1,3%

TÁBOA 12. Continuación

antes	1,3%
agora	1,3%
pode que	1,3%

Na lista anterior hai adverbios de foco como *só/soamente* (39a) e *tamén* (39b); aspectuais, como *sempre* (39c), *xa* (39d) e *aínda* (39e); enfatizadores, como *ben* (39f); de dúbida como a locución *pode que* (39g) e o de afirmación *si*, coa súa variante conxuncional *si que* (39h-i). Son proclisadores categóricos, é dicir, esixen de regra a anteposición:

- (39) a. Pero non só eles <o> requiren
 b. [...] senón que nos tamén <a> facemos no noso ámbito privado
 c. [...] os gobernos españois sempre <se> caracterizaron por seren corruptos
 d. [...] é decir, dende que nacemos xa <nos> creamos nunha sociedade consumista
 e. [...] á cociña aínda <lle> queda moito por explotar
 f. [...] na actualidade, como ben <se> di no texto, todos somos consumidores
 g. [...] e esté máis «calmado» pode que <os> entenda
 h. O correcto, que no caso de Carmela si <o> é
 i. Hai cousas máis importantes? Si que <as> hai

Pola contra, adverbios como *así*, *probablemente* e *incluso* (40a-d) (e outros como *antes*, *agora* e outros de modalidade en *-mente*) só provocan a próclise se van focalizados; se son Tópico, o pronome vai enclítico, como se viu no caso de *incluso* en (21g):

- (40) a. E así <se> crea un pilar caprichoso e pouco moral
 b. [...] probablemente a filla de Don Mariano <lle> pedira diñeiro para roupa, non ocorre ou non
 c. [...] moitos cultivaban as súas propias hortalizas, ou incluso <as> cambiaban por outros produtos básicos
 d. [...] de roupa, a veces moi semellante e incluso <nos> parece pouco

5.1.4. Cuantificadores focalizados

Na Táboa 13, pode verse a lista de cuantificadores en posición preverbal que provocan próclise por seren interpretados como Foco:

TÁBOA 13. Cuantificadores focalizados que provocan próclise

Proclisadores cuantificadores	% de casos
todo	55%
máis	15%
canto máis	10%
moito(s), -a(s)	10%
tanto	5%
cada	5%

Sexan adverbios, determinantes ou pronomes (41a-h), son os causantes da próclise no 2,1 % dos casos canónicos como indutores únicos, apenas 20 exemplos:

- (41) a. [...] todo está relacionado, todo <se> produce e <se> consume
 b. [...] isto ten parte de moda, a todo o mundo <lle> gusta quedar ben cos demais
 c. [...] todos e todas <nos> vemos afectados por unha inmensa cantidade de publicidade
 d. [...] parece que canto máis compras máis popular <te> fas
 e. [...] disfrutar desto que podemos chamar innovación e que tanto <nos> gusta
 f. Moitas veces <o> comparan con xente que vive na pobreza
 g. Coido que neste mundo, a comida todavía é algo co que non se debe xogar, pois, a moitísima xente <lle> falta
 h. A parte máis fácil de recoñecer e a do consumidor, cada cidadán <o> é

A regularidade de estes cuantificadores seren proclisadores non é tan alta coma a dalgúns dos adverbios comentados en 5.1.3 (*sempre, só, tamén, xa...*). Así, por exemplo, *todo* tende a provocar a próclise, pero aínda así atopamos algún exemplo xa comentado en (21d), no que a frase *todo o mundo* foi interpretado como Tópico e, de aí, a posposición do pronome; e o mesmo con *cada* en (42a-b) e *moitas veces* (42c), onde as frases cuantificadas por *cada* e *moitas* non formarían parte do mesma frase entoativa co clítico, polo que se produciría unha pausa ou inflexión antes do verbo e comezaría unha nova unidade fónica:

- (42) a. A cada un gústa<nos> o que se nos propón
 b. Cada vez complícan<se> máis os pratos e fan<se> máis exóticos
 c. Moitas veces nos costes de produción parecerían<nos> baratos

Véxase tamén o comentado nos exemplos de (21). A partir deles e do comentado neste apartado, poderíamos concluír que nos cuantificadores que

permiten unha lectura focalizada (próclise) ou a topicalizada (énclise) esta parece ser a opción preferente nos textos de *CORTEGAL*, redundando un tanto na tendencia á sobrexeneralización da énclise que tamén se observa noutros contextos de próclise habitual.

5.1.5. Outros constituíntes focalizados

Recollemos un exemplo (43) de frase focalizada (en maiúscula) como indutora única de próclise:

(43) NESO <se> basa o noso consumo

En (43), temos sintacticamente un complemento de réxime e informativamente un Foco contrastivo: entendemos ‘baséase NISO e non noutra cousa’.

Certamente, a tipoloxía textual de *CORTEGAL* non se axeita para procesos de focalización, como si o farían, por exemplo, discursos de conversa coloquial máis espontáneos. A focalización de constituíntes si é frecuente como indutor secundario, por exemplo cando hai xa outro proclisador primario, coma o nexo *que* en (44a-d):

- (44) a. É por iso que O CONSUMISMO <se> fixo de vital relevancia nas vidas diarias
 b. [...] porque os nosos país saben que ELES <o> pasaron aínda que en diferentes tempos
 c. [...] non é ningún misterio que A TODOS <nos> gusta comer, e sempre que a comida sea boa
 d. [...] aumentou o número de adeptos que co paso do tempo polo menos ALGÚNS <a> abandonarán, mais non por completo

5.2. Coordinación entre subordinadas

Nas estruturas coordinadas de oracións independentes, como xa vimos en (11a), o pronome sitúase enclítico, de non ser que se dean algunhas das condicións favorecedoras da próclise que vimos comentando. Non obstante, cando se coordinan dúas oracións subordinadas, caben dúas posibilidades, segundo Álvarez e Xove (2002, p. 564): (i) que os clíticos de todas as oracións coordinadas vaian en posición preverbal, opción preferible e obrigatoria se se repite o nexo; ou (ii) que só vaia proclítico no primeiro membro e enclítico nos seguintes, por causa de que se perda a idea de subordinación, posibilidade frecuente na oralidade.

O comportamento descrito axústase, polo xeral, ó rexistrado na mostra de *CORTEGAL*. Nos 11 casos en que o nexo se repite, o pronome vai proclítico

en 8 deles (45a-b); pero posponse en 3, perdendo a referencia do proclisador *que* por mor da distancia en (45c-d) e, agramaticalmente, cos dous nexos interrogativos de (45e):

- (45) a. [...] iniciar un ciclo que lle dé traballo e ademáis que <lle> garantice a legalidade laboral
 b. [...] aquí en Galicia gústanos comer ben e que <se> sirvan grandes racións
 c. [...] pensar que non sempre o fácil é o mellor camiño e que moitas veces o camiño corto manda <che> despois a o largo
 d. [...] (pode dicirse) ou que grazas a esta sociedade igualitaria máis xente interesou <se> por ela
 e. [...] podemos descubrir que diamante escóndese dentro del, e qué maravillas pódense ocorrer con él

No caso das coordinacións en que non se repite o nexo, dos 24 casos rexistrados, en 16 a posición é proclítica: o caso da relativa (46a), completiva (46b), temporal (46c), final (46d) ou condicional (46e). Non obstante, en 8 casos a posición rexistrada é a énlise, como vemos na temporal (46f), na completiva (46g), na causal (46h) ou na relativa (46i):

- (46) a. [...] a cociña era algo que se tiña que facer e <se> facía
 b. [...] é posible que fose collendo forzas e <as> manteña e perdure no tempo
 c. [...] plato de comida, mais cando o tes diante e <o> probas das te de conta de que vale a pena
 d. [...] para que as persoas e sobre todo os rapaces máis novos se interesen polo mundo da gastronomía e <se> integren nel
 e. [...] Se vives, <te> vistes, comes e traballas favoreces estas empresas
 f. [...] e dáste conta cando maduras e entéras<te> do que é a vida
 g. [...] E isto se debe a que se integrou na sociedade actual e encheu<se> de fama
 h. [...] porque a xente considera comer un placer e gústa<lle> buscar sabores novos
 i. Tamén están os restaurantes onde pagas moitos cartos e traen<che> un prato case vacío

En (46f-i), o segundo membro da coordinada non só xa perdeu o valor do nexo do membro 1, senón que a relación significativa entre os membros é xa tamén doutro tipo, manifestada agora través da conxunción copulativa e que toma novos valores: en (46f-g) de tipo consecutivo ('por tanto') e en (46h-i) de contigüidade.

6. CASOS DE ÉNCLISES E PRÓCLISES NON CANÓNICAS

Neste apartado rexistramos as colocacións que clasificamos como non canónicas ou agramaticais en galego, no senso de que non se axeitan ás regras xerais descritas nas gramáticas galegas, e que, polo xeral, son as observadas na lingua escrita estándar.

Detectamos na mostra analizada anomalías de dous tipos: casos de énclise cando o gramatical en galego é á próclise, coma o detectado en (8a); e o caso contrario, de próclise cando a regra xeral é a énclise, coma o que sucede en (8b)⁵, que reproducimos aquí de novo:

- (8) a. [...] este factor é determinante, xa que ninguén plantéxa <se> o feo que pode ser algo
 b. Estas dúas accións <as> leva facendo o ser humano case dende a súa aparición

O caso (8a) resulta agramatical xa que o indefinido negativo *ninguén* provoca a próclise do pronome (*ninguén se plantexa...*); en (8b), a anomalía radica en que estamos ante unha oración afirmativa e o pronome átono debe ir posposto (*Estas dúas accións lévaas facendo...*).

Como xa sinalamos na Táboa 5, detectamos un 5,2% de malas colocacións pronominais, unha porcentaxe que consideramos pouco elevada. Son máis frecuentes as malas prácticas de énclise por próclise (58,5%) do que as de próclise por énclise (40,3%). Nun único caso deuse asemade énclise e próclise, como vai dito xa, no exemplo (6).

Coidamos que estas anomalías poden ser achacables a dúas situacións que entremedian: por unha banda, a unha deficiente aprendizaxe (ou adquisición) do galego por parte do alumnado e, por outra, ó efecto do contacto lingüístico castelán-galego polo perfil bilingüe (aínda que sexa pasivo) maioritario no estudantado galego de bacharelato. Casos coma o de (8b) poden vir dados por un efecto da interferencia gramatical do sistema de colocación castelán (que presenta próclise nesas contextos) no do galego (que nesas situacións esixe énclise). Pola súa parte, anomalías coma a de (8a) non son efecto da interferencia gramatical do castelán (nesas contextos nas dúas linguas é de regra a próclise) senón do escaso dominio das regras de colocación, o que pode provocar unha ultracorrección a favor da posición enclítica, a opción non marcada no galego. No portugués europeo, cun sistema de colocación semellante ó galego, Brito et al. (2003, pp. 850-852) sinalan que durante a

⁵ Martins (2016, p. 410) tamén constata, con respecto ó portugués europeo padrón, casos de variación de próclise en contextos de énclise, e á inversa, en variedades africanas portuguesas e no portugués dialectal.

adquisición da lingua, as crianzas portuguesas «exibem generalizadamente o padrão enclítico, contrariamente ao que está descrito para as outras línguas románicas», e tamén que «as gerações máis jovens tendem a producir crescentemente clíticos enclíticos, en contexto en que a variedade padrão exige próclise». Achegan algúns exemplos como «(9)(a) porque não apercebeu-se que... (12 anos, modo escrito)» ou «(b) correspondem à classe onde «só» combina-se com SN... (estudante universitario modo escrito)», e conclúen que a «ênclise é o padrão de colocação em expansão nesta variedade». Costa et al. (2016, p. 441) achegan máis exemplos de énclices por próclises no portugués europeo e non só en crianzas senón tamén en adultos: «curiosamente, mesmo no grupo de controlo adulto, há 10,83% de produção de ênclice em contexto proclítico». Estes exemplos de colocación do portugués seméllanse a algúns dos rexistrados por nós, como o dito (8a) e outros que se comentan nos apartados 6 e 7.

Os casos de malas colocacións que localizamos na nosa mostra en alumnado con varios cursos de galego xa realizados deberan obrigar a facer unha reflexión sobre a docencia do proceso ensino-aprendizaxe neste aspecto da colocación pronominal e, de ser preciso, revisar ou reformular a súa didáctica así como a súa programación nos currículos escolares.

6.1. Casos anómalos de énclice por próclise

Os casos que rexistramos distribúense nos contextos que se sinalan na Táboa 14. Ordenámolos segundo a porcentaxe de casos anómalos con respecto ó total de exemplos analizados para cada contexto:

TÁBOA 14. Casos de colocacións anómalas de énclice por próclise

Contextos de énclice en lugar de próclise	% de aparición respecto ó seu total	% de casos anómalos
Con proclisadores de negación preverbais	1,3%	6,3%
Con adverbios focalizados preverbais	4,5%	8,3%
Con nexos preverbais	5,1%	70,8%
Con cuantificadores focalizados preverbais	26%	14,5%

A porcentaxe de aparición da anomalía podemos relacionala co grao do dominio das regras de colocación. Dende este punto de vista, os contextos negativos son os menos propensos a erros de colocación (1,3% de énclices nos casos analizados), logo virían os cometidos con adverbios preverbais focalizados (4,5% de énclices), seguidamente as estruturas introducidas por

nexos (5,1%), en tanto que os contextos cun cuantificador preverbal serían os máis inseguros e, daquela, nos que hai máis probabilidade de anomalías (26%); didacticamente, é os que máis conviría programar para traballar. A escala que nos resulta non difire moito da establecida por Costa et al. (2016, p. 444) no portugués europeo a partir dos seus estudos sobre as dificultades de adquisición da colocación de clíticos con crianzas (5-7 anos) e adultos. Eles propoñen a seguinte escala de desenvolvemento da próclise: *negación > suxeitos negativos/oracións completivas finitas > adverbio já > oracións adverbiais finitas > suxeitos cuantificados*, isto é, os primeiros contextos é onde antes se adquire a próclise; os últimos, onde máis tardiamente.

Debullamos máis os datos en cada contexto, indo do que ofrece menos anomalías porcentualmente ó que máis.

6.1.1. Con proclisadores negativos

Neste contexto rexistramos 3 énclices que deberan ser próclises por iren en oracións negativas: 2 casos co proclisador de negación *non* (47a-b) e o caso xa comentado do pronome indefinido *ninguén* de (8a). É unha colocación anómala moi pouco frecuente, pois só se deu no 1,3% dos casos analizados:

- (47) a. [...] fama dos grandes futbolistas, e non dámos<nos> de conta que os verdadeiros modelos...
 b. [...] que nin o paso de o tempo va<nos> facer olvidar o noso gusto pola comida

6.1.2. Con adverbios focalizados

Tamén resultan anómalos 4 exemplos de énclice con adverbios focalizados. Son os exemplos (48a-d) cos adverbios de foco *tamén* e *só*, o negativo *tampouco* e o de tempo *pronto*, pois son proclisadores categóricos. No caso de *pronto* mesmo vai precedido dun *que* interpretable como completivo (dependente de *penso*) ou mesmo como dos chamados «continuativos»; neste último caso, axudaría a inducir a próclise. As anomalías neste contexto supoñen un 4,5% con respecto ó total de adverbios focalizados:

- (48) a. Finalmente esta popularidade tamen debe<se> á televisión
 b. Solo coas opinións e coas críticas póde<se> mellorar e avanzar
 c. [...] á xente tampouco interésa<lle> moralmente, e prefiren elixir marcas baratas
 d. Por iso penso que isto da cociña trátase de algo pasaxeiro, que pronto esquecerémo<nos> de esta moda como nos esquecemos de moitas outras

O (48b) podería servir de exemplo de énclice por ultracorrección, xa que no texto (CORT_0068) o autor ou autora semella que primeiro redactou se

pode, riscou e corrixiu para *pódese*. Non obstante, mesmo neste exemplo, tamén puido facer durante a súa redacción unha pausa ou inflexión tras *críticas* e daquela a énclice sería posible ó se interpretar a secuencia *Solo coas opinións e coas críticas* como topicalizada.

6.1.3. Con nexos

Neste grupo de casos clasificamos a mala práctica de colocar o pronome proclítico habendo un nexo (conxunción, relativo ou interrogativo) en posición preverbal. Suman un total de 34 exemplos e a porcentaxe de erro con respecto ó total do seu contexto é de 5,1%. Na Táboa 15 recolleemos a tipoloxía, os nexos e o número de casos anómalos de énclice:

TÁBOA 15. Tipoloxía dos nexos que non provocaron próclise

Nexos	% de casos anómalos de énclice
Relativos <i>que</i> (7) <i>os que</i> (1) <i>a cal</i> (1) <i>onde</i> (1)	29,4%
Causais <i>porque</i> (3) <i>xa que</i> (3) <i>debido a que</i> (1)	20,6%
Completivos <i>que</i>	11,8%
Temporais <i>cando</i> (3) <i>mentres</i> (1)	11,8%
Finais <i>para que</i> (3) <i>co fin de que</i> (1)	8,8%
Interrogativos <i>que</i> (1) <i>como</i> (1) <i>cando</i> (1)	8,8%
Condicionais <i>se</i> (2)	5,9%
Modal <i>como</i> (1)	2,9%

A seguir listamos algúns exemplos de colocacións enclíticas cando hai un nexo, non permitidas no galego estándar. Os máis numerosos, como se ve na

Táboa 15, detéctanse con nexos relativos (49a-d), logo cos causais (50a-d), completivos (51a-c), temporais (52a-b), finais (53a-c), interrogativos (54a-b), condicionais (55a-b) e modais (56a):⁶:

- (49) a. As cousas que inténtan<nos> vender como a creatividade
 b. Son estes ambientes os que déixan<nos> situacións tan amargas
 c. [...] marca Toyypes, a cal gustába<lle> máis, isto fai que a marca Toyypes presente un ERE
 d. En conclusión, a gastronomía e un mundo onde quéda<nos> moito camiño por percorrer
- (50) a. Os compradores non pensan en aforrar eses cartos que se gastan na prenda, porque fai<lles> moita falta
 b. [...] impoñen castigos polo noso ben e porque preocupan<se> e queren o mellor para nos
 c. [...] do consumo, xa que produce <se> segundo a demanda
 d. Eu tamén teño problemas, debido a que fálta<me> conversas con eles
- (51) a. [...] o certo é que ségue<a> a haber
 b. Últimamente debo dicir, que dá<selle> maior difusión
 c. Que preguntas<lle> a un neno ou unha nena que queren ser
- (52) a. [...] pero cando vas<te> facendo maior dáste conta que esta situación pode producir conflitos
 b. [...] sentarse nunha terraza en sillóns masaxeadores mentres tráen<che> comida á carta
- (53) a. [...] enfocados á poboación co fin de que animémos<nos> máis a cociñar
 b. [...] nenos que queren ser de maiores para que un díga<che> astronauta
 c. [...] suficientes para que unha persoa aliménte<se> sen ter que deixar toda a súa nómina
- (54) a. [...] cando penso en como compórtaba<me> pero aprendín dos meus erros
 b. Cando este modelo de sociedade estoupará<nos> nas mans?
- (55) a. Si un bo gastrónomo gústa<lle> os pratos de o teu restaurante traeráche máis clientera
 b. Se dámos<lle> unha vida determinada a todos os productos, voume asegurar que vas volver compralos antes o despois
- (56) a. [...] prezo máis competitivo (como di<nos> o texto, Pascual fronte a unha «marca branca»)

Na maioría destes casos anómalos de énclice, entre o nexo e o verbo non hai ningún elemento. Si o hai nas finais de subxuntivo (53b-c), modo que por outra parte favorece a próclise, ou nas interrogativas indirecta e directa (54a-b). Anotamos isto porque, no caso dalgunhas relativas, completivas, causais e outras, rexistramos en *CORTEGAL* maiores porcentaxes de énclice do clítico cando se interpoñen outros constituíntes entre o nexo e o verbo. Son exemplos que comentamos no apartado 7, pois non os consideramos

⁶ Para un contraste de énclice e próclise en cada tipo de subordinada, véxanse logo os datos e comentarios das Táboas 19 e 20.

«agramaticais». Eses exemplos, e mesmo algúns dos incluídos entre (49) e (56), teríamos que cualificarlos máis como (?) ca (*). Por exemplo, nas subordinadas causais (50a-b) ou na completiva (51a), na lingua estándar escrita a próclise é o xeral, pero na oralidade a énclise é tamén unha opción. As gramáticas galegas chaman a atención sobre estes casos de énclise en subordinadas. Álvarez et al. (1986, p. 195) indican que «se lle debe dar preferencia á próclise, segundo a regra xeral que rexe para o pronome nas cláusulas dependentes». Pola súa parte, Álvarez e Xove (2002, pp. 565-566) cualifican estas posibilidades de énclise como de «estilo coloquial» e sinalan a próclise «máis recomendable». Freixeiro Mato (2000, p. 167) recolle a posibilidade, porén non avalía o seu uso para o estándar; limitáase a xustificar a énclise pola acción de posibles pausas e do comportamento variable dalgunhas conxuncións.

Énclises por próclises en subordinadas son atestadas tamén no portugués europeo por Brito et al. (2003, pp. 850-852), como xa vimos. Igualmente, Martins (2013, pp. 2276-2277) subliña a posibilidade de énclise en completivas, consecutivas e clivadas, aínda que cunha frecuencia moi reducida; e Ribeiro (2010, p. 24), tamén en temporais e modais. Pola súa parte, Fernández Lorences (2009, p. 95) documenta no asturiano «casos aparentemente anómalos en los que los clíticos, pese a coexistir con un transpositor, aparecen pospuestos al verbo», que son xustificados en ALLA (2000, p. 365) «pola existencia d'una pausa manifiesta despues de *que*. Exemplos: *ye que me molestaba enforma* o *ye que molestábame enforma*».

6.1.4. *Con cuantificadores focalizados*

O último grupo de exemplos cunha colocación enclítica anómala rexistrámolo cando en posición preverbal hai un cuantificador focalizado. Anotamos 7 casos de *todo*, dos que exemplificamos algún en (57a-d). É o contexto de colocación en que se observan máis anomalías, un 26% con respecto ó total dos casos de cuantificadores tidos en conta:

- (57) a. [...] iso nunca cambiará, a todos gústa<nos> unha boa tortilla de patacas ou lacón con grelos
 b. A todos gústa<nos> comer ben, e se non o temos que cociñar
 c. Porque todo o mundo fala deles e todo o mundo probou<nos>
 d. [...] dous ou tres destes prestixiosos restaurantes e todos déron<me> a mesma impresión

Sobre a posibilidade de que algúns outros cuantificadores (e adverbios) preverbaais fosen interpretados como Foco (o que esixiría a próclise), véxase o dito nos exemplos de (21). Ás veces resulta dificultoso vulgar a gramatical-

dade do clítico con esta clase transversal de palabras, pois depende da entoación e énfase que se lle imprima e da interpretación como Focos ou Tópicos, coma tal o caso (57d). Desde o punto de vista didáctico, é este un aspecto pragmático que precisará de máis atención no proceso de ensino-aprendizaxe da colocación pronominal.

6.2. Casos anómalos de próclise por énclise

Recollemos neste apartado as malas colocacións proclíticas do pronome en casos en que deberan ir enclíticos ó verbo tratándose de oracións simples, principais de subordinadas e segundos membros de coordinadas copulativas de *e* e adversativas de *pero*. Nestes casos, infrínxese a regra de que en galego o pronome non pode ir en posición preverbal de non se dar algunhas das condicións xa ditas: presenza na periferia verbal dunha negación, dun nexo ou de constituíntes focalizados, sexan cuantificadores, adverbios ou calquera outro elemento.

Nos exemplos analizados non atopamos ningún caso que contradiga a lei Tobler-Mussafia, segundo a cal un clítico non pode iniciar nunca o enunciado. No resto de *CORTEGAL* é tamén unha anomalía moi pouco frecuente; apenas localizamos 2 exemplos con <se>: <Se> *debe ter presente á hora de producir*; <Se> *pode observar co simple feito en que a xente cando sale de traballar*.

Non obstante, si anotamos 33 exemplos doutras malas colocacións de próclise por énclise, o 40,3% de todas as agramaticais. En (58) listamos algúns exemplos de oracións principais e independentes con anteposicións pronominais sen haber ningún elemento proclisador. Son próclises non admitidas no galego estándar:

- (58) a. A evidencia <nos> di que o ser humano mellorou as súas condicións alimentarias
 b. Eu <o> considero un cambio cultural moi importante, enriquecedor e duradeiro
 c. O simple feito de que as nenas queren ser modelos e os nenos futbolistas, <nós> amosa esa desigualdad que nós debemos erradicar
 d. Estas dúas accións <as> leva facendo o ser humano case dende a súa aparición
 e. O enfoque mercantil <o> recibimos de maneira inconsciente por toda a sociedade
 f. Isto <o> fan para que a poboación non se pense tanto as cousas
 g. Eses cartos <os> sacamos traballando, grazas ao noso salario
 h. Moitas veces <nos> levamos polo bonito e non polo necesario
 i. Neste «novo período» <nos> atopamos nun momento de desenvolvemento e investigación de novos

- j. Para rematar, <me> gustaría engadir que outro medio capaz de resolver todo tipo de dúbidas no tema da gastronomía ou cociña é o máis coñecido: Internet
 k. E <me> pregunto eu, como é posibel poder obter tanta felicidade

Os constituíntes que inician a oración poden ser SUX (58a-c), CD redobrados (58d-g), complementos circunstanciais (CCIRC) (58h-i), un modificador oracional (58j) e a conxunción copulativa *e*, neste caso con valor de conector discursivo (58k). Agás este último caso, todos os outros os interpretamos informativamente como Temas ou Tópicos (agás o caso de 59k) e, daquela, non é posible a próclise. De feito, nos CD, pode indicarse que unha marca de que son Tema ou Tópico é o redobro do clítico (se fosen focalizados o dobrado non é posible).

Tampouco son correctas as próclises pronominais das oracións principais que van anticipadas polas súas subordinadas nos exemplos de (59). Son casos coma as das apódoses condicionais (59a-b), a tese da concesiva (59c) ou a matriz das temporais de (59d):

- (59) a. [...] se non producimos o suficiente, por exemplo <nos> despíden do traballo
 b. [...] porque hai que mercar a alí, se non, <te> botan fóra
 c. Se ben o recente éxito deste mundo foi moi espontáneo, a importancia que ten <lle> asegura unha larga estancia
 d. Cando ves a televisión, cando vas pola estrada e ves un cartel con publicidade, ou cando te poñen propaganda no buzón de casa, <te> estan suxestionando para que compres

E igualmente resultan agramaticais as próclises en oracións que conforman o segundo membro de estruturas xustapostas como (60a-b) e de adversativas como (60c-d):

- (60) a. [...] contradicir a educación que súa nai lle está impoñendo, <me> refiro a que si un lle consinte unha cousa...
 b. As clases produtoras son despreciadas; <as> tratamos como un medio
 c. [...] mais o Estado de bienestar e un alto nivel de vida para todos os cidadáns <se> logran cando non hai tanta diferenza entre uns salarios
 d. [...] porque lles gusta a comida, pero outros <o> fan por pura diversión e para vivir unha nova experiencia

7. OUTRAS COLOCACIÓNS DIVERXENTES DA POSICIÓN CANÓNICA

Incluimos aquí dous tipos de diverxencias con respecto á posición canónica: casos de énlises en subordinadas cando entre o nexa e o verbo hai outros constituíntes (65 casos) e casos de próclise en subordinadas tras pausa (9 casos). En conxunto supoñen o 4.7% das colocacións analizadas.

7.1. Casos de énclices en subordinadas

Xa sinalamos en §4.2 e §6.1.3 a posibilidade de que, nalgunhas oracións subordinadas con nexo, o pronome átono poida ir en posición enclítica, coma o xa visto (50a). Son casos que a lingua estándar escrita prescribe como proclíticos (50a'), pero a énclice (50a) pode ser tamén unha posibilidade, cando menos para algúns falantes no galego oral, e cumpriría, didacticamente, telo en conta, sobre todo para alumnado galego falante:

- (50) a. Os compradores non pensan en aforrar eses cartos que se gastan na prenda, porque fai <lles> moita falta
 a'. [...] eses cartos que se gastan na prenda, porque lles fai moita falta (*estándar escrito*)

Na posición enclítica, *porque failles moita falta*, o nexo causal e a forma verbal co clítico fan parte de frases entoativas diferentes: [porque] [failles moita falta] e, de aí, que o pronome, ó haber grupo fónico novo, se pospoña; pola contra, en *porque lles fai moita falta* nexo, clítico e verbo si pertencen á mesma frase entoativa [porque lles fai moita falta]. Ademais, notamos que énclice e próclise son informativamente diferentes: na énclice (*porque failles*) a causa enfatiza na acción de 'facer'; na próclise (*porque lles fai*), a causa fócuse nos destinatarios (*lles*). Mesmo tradicionalmente se ten apuntado que nos casos de énclice o nexo *porque* posúe valor explicativo, equivalente a 'pois', e nos de próclise é puramente causal.

Como xa se comentou naquelas epígrafes, as gramáticas galegas, recoñecendo que a próclise nas subordinadas é «de regra» (Álvarez et al., 1986, p. 194), reservan apartados nos que describen estas «excepcións» de énclices con oracións completivas, relativas, causais e consecutivas (Álvarez et al., 1986, p. 194; Freixeiro Mato, 2000, p. 166; Álvarez e Xove, 2002, pp. 565-656). Aluden á «permisibilidade» na posición do clítico e cualifican a énclice como de estilo máis coloquial e a próclise como «máis recomendable» (Álvarez e Xove, 2002, p. 566). Por outra parte, Ogando (1980, p. 266) no galego medieval e Meier (1976) nos contos populares lugueses do século pasado, documentaron énclices semellantes en subordinadas; e nós mesmos, como falantes, tamén afirmamos esa posibilidade no galego oral, poida que con matices de significado informativo coma os comentados arriba a propósito de (50a-a'). Igualmente, en portugués (Brito et al., 2003; Martins, 2013; Ribeiro, 2010, p. 24) e en asturiano (D'Andrés Díaz, 1993; Fernández Lorences, 2009), rexístrase esta posibilidade, aínda que non sexa a opción promovida para a lingua padrón. Polo tanto, estas posibilidades de colocación existen

no galego, e nas outras linguas con sistemas de colocación semellantes, outra cuestión é o status que se lle dea na lingua estándar.

As explicacións de énclise para estes casos inciden no valor dos nexos (matices explicativos, causais, consecutivos, *que* neutros, fórmulas introdutorias como *é que*, comportamentos de marcadores discursivos), en roturas entoativas entre nexo e clítico provocadas por pausas e na inclusión de constituíntes entre o nexo e o verbo, que ó seren interpretados como Tópico ou Tema provocan énclise.

Nesta epígrafe agrupamos precisamente este tipo de casos: subordinadas con nexos de diferente tipo, co pronome enclítico e con diferentes tipos de constituíntes oracionais situados entre o nexo e o verbo.

Na Táboa 16 poden verse clasificados polo tipo os 65 exemplos de maioritariamente subordinadas con secuencia NEXO + CONSTITUÍNTE + VERBO + CLÍTICO en que rexistramos énclise, así coma a súa porcentaxe de ocorrencia; nelas incluímos tamén 2 casos de antíteses adversativas (consideradas tradicionalmente coordinadas):

TÁBOA 16. Tipoloxía de oracións en que se dan casos de énclise pronominal cando hai un constituínte entre o nexo e o verbo

Tipo de oración	NEXO+CONSTIT.+VERBO+ CLÍTICO
Causal	33, 8%
Completiva	26,1%
Relativa	16,9%
Temporal	4,6%
Causal consecutiva	4,6%
Interrogativas	4,6%
Condicional	3,1%
Concesiva	3,1%
Antítese de adversativas	3,1%

Como se ve, as causais, completivas e relativas suman case o 77% do total, ben que destacan as causais con máis de 1/3 dos casos. Ofrecemos a seguir exemplos de énclise nas causais (61), completivas (62), relativas (63), temporais (64), causais consecutivas (65), interrogativas indirectas (66), condicionais (67), concesivas (68) e de antíteses adversativas (69):

- (61) a. [...] xa que o prezo dos servizos e produtos é<lles> fácilmente asumible
 b. [...] senón porque o simple feito de telo fai<nos> ser máis «populares»
 c. [...] posto que a nosa sociedade trata<se> dunha sociedade de consumo

- d. [...] dado que todo o sistema económico dun país susténta<se> en que as empresas produzan
- e. Xa que á outra persoa podería<lle> sentar mal
- f. [...] xa que todo o que vamos deixando atrás, tirámo<lo> e en moitos casos aun están para o uso
- g. [...] debido a que, hoxe en día, póde<se> chegar dun xeito máis directo ás persoas
- h. [...] xa que moitas veces para a produción de certos artigos precisa<se> de recursos naturais
- (62) a. Eu creo que isto debe<se> ao turismo
- b. [...] é certo que as persoas témos<nos> que adaptar á nosa condición económica
- c. [...] engadir que a gastronomía hai<na> tan asequible como para que a maior parte...
- d. [...] no feito de que no século xx levou<se> a cabo a terciarización
- e. Tamén que á xente gústa<lles> saber cociñar para aprender a facelo casa
- (63) a. Foi a curiosidade a que pouco a pouco fixo<nos> interesar máis pola comida
- b. [...] o resultado final (onde moitas veces cóida<se> máis a estética que o sabor)
- c. [...] e por isto que moitos adolescentes revélan<se>
- d. Galicia goza dunha comida que aos estranxeiros chába<lles> a atención
- e. [...] a cal, como xa dixen, permíte<me> ter este tipo de visión
- (64) a. O problema aparece cando ese consumo vólve<se> obsesión
- b. [...] ocorre cando o pai ou, neste caso a nai, di<lle> que non a algo
- c. [...] é que cando algo pon<se> moi de moda ti se non o tes, eres peor ca o resto
- (65) a. [...] iso non ocorre nestes países, polo que á comida póde<selle> engadir outros alicientes
- b. [...] comezase a engordar, polo que agora preténde<se> implantar unha dieta máis sana
- c. [...] polo que con sorte, mínimo toca<che> esperar 6 meses para poder ir
- (66) a. Resulta impresionante ver como do día a mañá en apenas vinte anos a gastronomía voltou<se> unha profesión grandiosa
- b. [...] segundo señora ve como a súa empresa vai<se> a pique pero finalmente ela compra a leite
- (67) a. [...] se a iso sumámos<lle> o cambio climático a situación pode chegar a ser catastrófica
- b. [...] para moitos é un gran pracer e se a eso engadímos<lle> a posibilidade de aprender forman unha combinación moi boa
- (68) a. [...] a pesar de que cada vez máis elaboran<se> comidas de mala calidade e moi procesadas
- b. [...] único de mulleres (aínda que os grandes premios léven<os> os homes sempre)
- (69) a. [...] parece exquisito pero que finalmente ofrécen<cho> mal cocinado
- b. [...] o de sempre non é o mellor, senón que no gusto e na diversidade atópa<se> o equilibrio

O constituíntes inseridos (que subliñamos nos exemplos) son, en máis da metade dos casos, sintacticamente SUX, boa parte deles longos e complexos,

como en (61a,b,d), pero tamén os hai curtos, como (62a). Seguen en frecuencia CCIRC e outros adxuntos periféricos, con case 1/4 dos casos, algúns tamén complexos (69b) e, en (66a), dous topicalizados. Rexistramos tamén CI (62e) ou (67a-b); CD redobrados (62c), (68b), ou o topicalizado (61f), e incisos, como (61g), (63e) ou (64b) (aínda que este mal puntuado). A énclice en boa parte destes exemplos, de dalos como posibles, cómpre explicala a partir de que estes constituíntes, informativamente, foron concibidos polo redactor ou redactora dos textos como Tópico ou Tema, aínda que a puntuación non sempre reflecta iso.

Obviamente, se estes elementos non se interpretan topicalizados, varias destas colocacións enclíticas poden resultar forzadas e mesmo ser cualificadas de non admisibles en galego, por exemplo (63a-b), (66b) ou (68b). Noutros casos, como (64a,c), (67a-b) ou (63c), se cadra o máis habitual sexa enuncialos focalizados e daquela o máis frecuente sexa a próclise, a versión preferida na lingua estándar escrita. Pola contra, algunhas énclices en casos de CI que adoitan ser Tópico ou Tema, como (61e) ou (63d), parécenos tamén totalmente habituais na oralidade e admisibles na escrita, como tamén o sería nesta a énclice nos incisos (61g) e (63e).

Na Táboa 17 constatamos as ocorrencias das distintas funcións dos constituíntes segundo a clase de subordinada.

TÁBOA 17. Funcións sintácticas dos constituíntes inseridos entre o nexa e o verbo en casos de énclice pronominal

Tipo de oración	SUX	CD	CI	CCIRs/ Adxunt.	CCIR- SUX	INC.
Causal	9	3	4	4		2
Completiva	10	2	1	3	1	
Relativa	5		1	4		1
Temporal	2					1
Causal consecutiva			1	2		
Interrogativa	2				1	
Condicional			2			
Concesiva		1		1		
Adversativa				1	1	
Total (%)	28 (43,1%)	6 (9,2%)	9 (13,8%)	15 (23,1%)	3 (4,6)	4 (6,1%)

Por outra parte, na Táboa 18 ordenamos por frecuencia os nexos que introducen as 65 oracións, maioritariamente subordinadas, en que se dá énclice:

TÁBOA 18. Ocorrencias de nexos que non provocan próclise

Nexos	Tipo	% de casos
que	completivo	26,2%
xa que	causal	17%
porque	causal	10,8%
que	relativo	10,8%
polo que	causal consecutivo	4,6%
como	interrogativo	4,6%
cando	temporal	4,6%
se	condicional	3,1%
o cal	relativo	3,1%
debido a que	causal	3,1%
senón que	adversativo	1,5%
pero que	adversativo	1,5%
posto que	causal	1,5%
dado que	causal	1,5%
aínda que	concesivo	1,5%
a pesar de que	concesivo	1,5%
como	relativo	1,5%
onde	relativo	1,5%

Os catro primeiros nexos, *que* completivo, *xa que* e *porque* causais e *que* relativo, suman case o 65% de todas as ocorrencias. Por outra parte, entre a lista tamén localizamos as locucións conxuntivas adversativas *pero que* e *senón que*, que malia incorporaren a conxunción *que* e a idea negativa no caso de *senón*, non actuaron de proclisadores. Casos de *senón que* recolleemos 4 exemplos na nosa mostra, en 3 esixía próclise (25a-c) e no outro o pronome ía posposto (69b).

Como conclusión, gustaríanos contrastar datos das táboas 9 e 16, isto é, casos de éclise e próclise en subordinadas con nexo en que tamén pode haber outros constituíntes en posición preverbal; e tamén datos da Táboa 15, de nexos que non provocaron próclise. Recóllense na Táboa 19:

TÁBOA 19. Próclise e énclise en oracións subordinadas con nexo

Tipo de subordinada	Próclises canónicas		Énclices non canónicas	
	NEXO+CLÍTICO+VERBO	NEXO+CONST.+CLÍTICO+VERBO	NEXO+VERBO+CLÍTICO	NEXO+CONST.+VERBO+CLÍTICO
Relativa	286	28	10	11
Completiva	71	38	4	17
Condicional	23	21	2	2
Causal	25	15	7	22
Temporal	31	5	4	3
Modal	20	-	1	-
Comparativa	19	-	-	-
Final	5	12	3	-
Concesiva	7	1	-	2
Consecutiva	1	1	-	
Causal consecutiva	1	1	-	3
Interrogativa indirecta	11	-	3	3
Total	500	121	34	63
	80,5%	19,5%	35,1 %	64,9%
	621		97	
	86,5%		13,5%	
	718			

Globalmente, nos exemplos analizados de *CORTEGAL* de estruturas subordinadas con nexo (718 casos), o habitual é a próclise no 86,5%; pero esta non é de regra, pois hai un 13,5% de énclices nas que o nexos non actúa de proclisador. Ademais, nos casos de énclise, esta é máis habitual, nun 64,9%, cando entre o nexos e o verbo se insiren outros constituíntes oracionais (e así a distancia entre o nexos e o clítico é maior e as posibilidades de quebra entoativa tamén); se non hai ningún elemento as posibilidades de énclise baixan ata un 35,1%.

Por outra parte, os casos de próclise e énclise non son iguais en todas as subordinadas. Téndomos en conta as estruturas con máis ocorrencias, interésannos aquelas en que se dan máis excepcións á regra. Así, segundo a Táboa 20, as porcentaxes máis altas de énclise en *CORTEGAL* (vaia o nexos apegado ó verbo ou non) danse nas causais (42%), temporais (16,3%),

completivas (16,2%) e relativas (6,2%)⁷, o que en xeral confirma o apuntado nas gramáticas galegas:

TÁBOA 20. Porcentaxes de próclise e énclise nas subordinadas con nexos máis numerosas

	Próclise	Énclise
Causais	58%	42%
Completivas	83,8%	16,2%
Temporais	83,7%	16,3%
Relativas	92,8%	6,2%

Nas subordinadas de menos ocorrencias, é salientable que nas que clasificamos como causais consecutivas de *polo que* (65) hai énclise nun 60%. Pola contra, parece de regra total a próclise nas comparativas, nas modais nun 95,2%, nas finais nun 80% e un 64,7% nas interrogativas indirectas.

E de seguirmos debullando máis nos datos da Táboa 19, poderíamos dicir que nas causais, as subordinadas máis permeables á énclise, se hai un constituínte entre o nexos e o verbo, a probabilidade de que o pronome vaia enclítico é na mostra de *CORTEGAL* do 59,4%; nas completivas, do 30,9%; nas temporais, do 37,5%; e nas relativas, do 28,2%. No caso de non haber constituínte, a porcentaxe de énclise baixa considerablemente: aínda así, nas causais segue a ser dun 21,9%, pero nas temporais é xa dun 11,4% e moito menos nas completivas, 5,3%, e relativas, 3,4%.

Como conclusión a este apartado, pódese dicir que, no alumnado autor dos textos analizados, a énclise, aínda sendo minoritaria no conxunto das subordinadas con nexos (13,8%), ben seguro que sería porcentualmente moito maior cá que se podería detectar nun corpus de galego escrito da variedade estándar. Non obstante, as porcentaxes de posposición do pronome átono son considerables en estruturas causais nas que hai elementos interpostos entre o nexos e o verbo. Esta tendencia á énclise, na parte do alumnado que sexa galegofalante, debe ser explicada internamente (dentro do propio galego), pois nestes contextos a posposición é unha posibilidade constatable (e tamén no portugués e no asturiano) e, á hora de se ensinar, isto debera terse en conta. No alumnado non galegofalante habitual, as causas de énclices como as comentadas nos exemplos de (61) a (69) deberán buscarse nun incompleto proceso de ensino-adquisición dos mecanismos que rexen

⁷ No estudo xa citado, Costa et al. (2016, p. 441) documentan en adultos portugueses un 22,4% de énclices no contexto de «oracións adverbiais» (causais, temporais, consecutivas...) e 2,3% nas completivas.

a colocación neste contexto: isto crearíalle a inseguridade nun sistema de colocación galego que xa de seu é complexo e, ante a dúbida, este alumnado optaría, quizais por ultracorrección, a sobrexeneralizar a posición enclítica, a non marcada en galego.

7.2. Casos de próclise en subordinadas tras pausas gráficas

Agrupamos aquí un feixe de 9 exemplos que teñen en común presentaren o pronome átono proclítico tras unha pausa, representada por coma (,) ou paréntese, aínda que a puntuación non sempre é a correcta. Son os casos da causal (70a), a causal consecutiva (70b), as dúas completivas (70c-d) e as cinco relativas (70e-i):

- (70) a. [...] mentres que un grupo máis reducido, <che> diría que prefire salvar vidas nun hospital
 b. [...] polo que, a menudo, <os> fan escoller a que consideran correcta
 c. [...] a mocidade cree que teñer un corpo bonito (aínda que non posea coñecementos dentro), <lles> solucionará a vida
 d. Como moitas veces pasa nas familias, que os fillos sen ter culpa, <as> pagan igualmente
 e. [...] en numerosas empresas que, por exemplo, grazas á publicidade ou outros aspectos de marketing <nos> incitan a mercar e consumir moitos produtos
 f. Esta idea vese alterada nos anuncios de beleza, onde, a frase «quérete a ti mesmo» <che> ensina a un home ou muller perfectos
 g. Coas que en lugar de telo control do noso diñeiro, <o> entregamos ás empresas
 h. [...] por tanto unha clase social maior, agás excepcións que, como ben din o autor no texto, <as> hai para calquera regra
 i. [...] de obxectos, que aínda xa tendoos, <os> volvemos a comprar porque son unha versión superior

Por exemplo, en (70a) se facemos a pausa onde marca a coma (,) implicaría que o que hai antes pertence a outra frase entoativa e daquela o pronome cliticaríase posposto ó verbo: o constituínte *un grupo máis reducido* interpretaríase informativamente como Tópico/Tema ('mentres que, canto a un grupo máis reducido, diríache que...'). Así e todo, tamén cabería interpretar que o redactor ou redactora do texto fixese un uso deficiente da coma. Eliminándoa, e con ela a pausa na lectura, a próclise sería posible, aínda que se podería explicar de dous xeitos: 1) interpretar *un grupo máis reducido* como un SUX Tema e a próclise viría xustificada pola acción do nexo temporal; ou 2) interpretalo como un SUX focalizado, contrastivo, enfatizado de xeito que se puidese interpretar 'mentres que UN GRUPO MÁIS REDUCIDO (e non moita xente) che diría...'; tamén axudaría a influencia do nexo temporal.

Na nosa opinión, nos casos de (70) a énclice sería preferible á próclise, pois as pausas e inflexións (representadas ou non) que cómpre facer oralmente obrigarían a unha rotura entoativa e reiniciar a oración co pronome posposto. Non obstante, Álvarez e Xove (2002, pp. 565-566) sinalan para estes casos de incisos que «nun rexistro coidado tende a evitarse este cambio de posición».

8. CONCLUSIÓNS

Neste traballo analizamos a posición de 1578 pronomes átonos con respecto a formas verbais finitas do corpus *CORTEGAL*. O noso obxectivo foi ofrecer unha visión global da colocación pronominal átona en textos escritos producidos por alumnado galego de bacharelato. Os datos obtidos foron clasificados e comentados en contraste coa posición canónica que se observa no galego estándar escrito e que é a recollida nas gramáticas galegas. Tamén se contrastaron con outros traballos cando existían e cos doutras linguas coma o portugués ou asturiano que, en xeral, presentan contextos de énclice e próclise pronominal semellantes ás do galego.

A tipoloxía textual de *CORTEGAL* determinou que na mostra investigada os casos de próclise sexan maioría nun 64,8% fronte ós de énclice, o 35,1%. As colocacións pronominais examinadas son coincidentes coas do galego estándar nun 90,1% dos casos e só nun 9,9% se produce algunha diverxencia: unha porcentaxe pequena, ó noso ver. Dentro dela, un pouco máis da metade (5,2%) considerámolas non admisibles no galego estándar escrito e, didacticamente, será nestes casos nos que os docentes deberán insistir máis; nas restantes, un 4,1% son énclices posibles no galego oral, ben que non son as preferibles na variante escrita culta; finalmente, un 0,6% resultan casos raros de próclise cunha pausa antes marcada por algún signo de puntuación.

A énclice pronominal é a opción non marcada en galego e, como tal, é a que se manifesta nos casos que consideramos de ultracorrección; son contextos que, metodoloxicamente, tamén precisarían de atención especial no proceso de ensino-aprendizaxe deste aspecto. Nos casos de colocación canónica, constatamos que o pronome se despraza para antes do verbo, sobre todo, pola acción de nexos relativos e conxuncionais (un 65,6% dos casos) e de proclisadores de negación (23,2%); e en menor medida, por mor da presenza de elementos focalizados en posición preverbal, sobre todo adverbios (8,9%) e cuantificadores (2,1%).

Nos casos de colocación non canónica, as anomalías son máis numerosas en colocacións enclíticas que deberan ser proclíticas, un 58,5% (**este factor*

é determinante, xa que ninguén plantéxa <se> o feo que pode ser algo > ... ninguén se plantexa...), do que de próclises por énclices, un 40,3% (*Estas dúas accións <as> leva facendo o ser humano case dende a súa aparición > Estas dúas accións lévaas facendo...). As primeiras aventuramos explicalas por unha mala aprendizaxe do galego como L2 —o que debera servir ós docentes e programadores de cursos para facer unha reflexión sobre a pedagogía seguida neste aspecto da gramática galega—; e as segundas, por efecto da interferencia lingüística do castelán; pero as dúas causas tamén se entrecruzan. Con todo, non en todos os contextos a probabilidade de anomalía de énclice por próclise é a mesma: polos nosos datos, no contexto negativo é onde menos se producen porcentualmente e o dos cuantificadores focalizados onde máis.

Nunha epígrafe á parte comentamos outros casos diverxentes: os de próclise despois de pausa (*Como moitas veces pasa nas familias, que os fillos sen ter culpa, <as> pagan igualmente*) —que, ó noso parecer, irían mellor enclíticos—, e os máis numerosos e controvertidos de oracións subordinadas con nexos en que este non actúa de proclisador. Neles, a posposición do clítico constatouse nun considerable 13,5% do total de subordinadas, aínda que as porcentaxes varían dependendo de se entre o nexos e o verbo entremedian ou non outros constituíntes oracionais. Nos casos de énclice, de o nexos e o verbo iren contiguos (*o certo é que séguese <a> a haber*) a énclice dáse nun 35,1% de exemplos; son colocacións non admitidas na variante estándar. De haber outros elementos (*xa que o prezo dos servizos e produtos é <lles> fácilmente asumible*), a porcentaxe de posposición pronominal sobe ata o 64,9% dos casos: son énclices posibles no galego oral, pero non recomendables no galego estándar escrito segundo as gramáticas galegas. As causais, completivas, temporais e relativas son os tipos onde máis documentamos estas énclices; mesmo nas causais, os datos sinalan que se hai un constituínte entre o nexos e o verbo, a probabilidade de que o pronome vaia enclítico é na mostra examinada de CORTEGAL dun 59,4%.

Nestes casos, tamén cos adverbios, cuantificadores ou calquera tipo de frases en posición preverbal, para avaliar a colocación do clítico como lexítima ou non no galego é fundamental a súa interpretación informativa: se os topicalizamos favorecemos a énclice; se os focalizamos, a próclise.

Finalmente, queremos salientar a utilidade de CORTEGAL como ferramenta para avaliar o proceso de ensino-aprendizaxe da lingua galega. Na mostra analizada, chamamos a atención sobre aqueles contextos en que se producen anomalías ou diverxencias na colocación do pronome con respecto

á posición canónica do galego estándar escrito. Como liñas de investigación futuras da nosa análise, é conveniente indagar sobre as causas destas diverxencias en alumnado que leva varios anos cursando galego. Sería esencial ter en conta o perfil (socio)lingüístico do estudantado, pero tamén examinar a didáctica empregada no ensino deste aspecto gramatical, revisar o currículo e ver en que cursos se introducen os pronomes, cando se estudan os clíticos e durante cantos cursos, as dificultades que este aspecto supón nos aprendentes, as necesidades de reforzalo e mesmo cal sería a metodoloxía máis eficiente. Tamén nos parece fundamental ter en conta a variación diamésica, pois nalgúns contextos, como as oracións subordinadas con nexos, a colocación pode diverxer do galego oral ó galego escrito. Ademais, coidamos que as diverxencias e anomalías de colocación detectadas non se deberan abordar unicamente na clase de lingua senón tamén naquelas outras materias do currículo que teñen o galego como lingua vehicular.

REFERENCIAS BIBLIOGRÁFICAS

- ALLA = Academia de la Llingua Asturiana (2001). *Gramática de la Llingua Asturiana*. Academia de la Llingua Asturiana. <http://www.academiadelalingua.com/wp-content/uploads/2014/10/Gramatica-Llingua.pdf>
- ÁLVAREZ BLANCO, M. R. (1980). *O pronome persoal en galego* [Tese de doutoramento inédita]. Universidade de Santiago de Compostela.
- ÁLVAREZ, R. e XOVE, X. (2002). *Gramática da lingua galega*. Galaxia.
- ÁLVAREZ, R. (1996). Posición do pronome átono en construcións de infinitivo e xerundio. *Verba*, 23, 11-35. <http://hdl.handle.net/10347/3264>
- ÁLVAREZ, R., REGUEIRA, X. L. e MONTEAGUDO, H. (1986). *Gramática galega*. Galaxia.
- BRITO, A. M., DUARTE, I. e MATOS, G. (2003). Tipología e distribución das expresións nominais. En M. H. M. Mateus, A. M. Brito, I. Duarte e I. Hub Faria, *Gramática da Lingua Portuguesa* (5.^a ed. rev. e aum, pp. 795-867). Caminho.
- COSTA, J., FIÉIS, A. e LOBO, M. (2016). A adquisición dos pronomes clíticos no português L1. En A. M. Martins e E. Carrilho (Eds.), *Manual de Lingüística Portuguesa* (pp. 431-452). De Gruyter. <https://doi.org/10.1515/9783110368840-018>
- D'ANDRÉS = ANDRÉS DÍAZ, R. d' (1993). *Allungamiento de los pronomes átonos col verbu n'asturainu*. Universidá d'Uviéu, Departamento de Filología Española.
- FERNÁNDEZ LORENCES, T. (2009). Ámbito del predicado y posición de los pronombres átonos en asturiano. *Revista de Filoloxía Asturiana*, 9(9-10), 87-111. <https://reunido.uniovi.es/index.php/RFA/article/view/9176>

- FREIXEIRO MATO, X. R. (2000). *Gramática da lingua galega. II. Morfosintaxe*. A Nosa Terra.
- FREIXEIRO MATO, X. R. (2003). *Gramática da lingua galega. IV. Gramática do texto*. A Nosa Terra.
- GONZÁLEZ I PLANAS, F. (2007). Sintaxis de los clíticos pronominales en asturleonés. *Ianua Revista Philologica Romanica*, 7, 15-35. <http://romaniaminor.org/ianua/Ianua07/02.pdf>
- MARTINS, A. M. (2013). A colocação dos pronomes pessoais clíticos. En E. B. Paiva Raposo, M. F. Bacelar do Nascimento, M. A. Coelho da Mota, L. Segura e A. Mendes (Orgs.), *Gramática do Português* (pp. 2231-2302, vol. II). Fundação Calouste Gulbenkian.
- MARTINS, A. M. (2016). A colocação dos pronomes clíticos em sincronia e diacronia. En A. M. Martins e E. Carrilho (Eds.), *Manual de Lingüística Portuguesa* (pp. 401-430). De Gruyter. <https://doi.org/10.1515/9783110368840-017>
- MEIER, H. (1976). Problemas de gramática galega. *Verba*, 3, 45-52.
- OGANDO VÁZQUEZ, V. (1980). A colocación do pronome átono con relación ó verbo no galego-portugués. *Verba*, 7, 251-282.
- PRIETO ALONSO, D. (1986). Movimentos dos clíticos em galego-português. En *Actas do I Congreso Internacional da Língua Galego-Portuguesa na Galiza (Ourense, 1984)* (pp. 359-366). AGAL.
- RIBEIRO, I. (2010). Sobre os usos de ênclise nas estruturas subordinadas no português arcaico (About the use of enclisis on subordinated structures in Old Portuguese). *Estudos da Língua(gem)*, 8(1), 15-40. <https://doi.org/10.22481/el.v8i1.1113>
- SACO ARCE, J. A. (1868). *Gramática gallega*. Imprenta de Soto Freire.
- SANTOS RAÑA, I. (2018). Variación na colocación dos pronomes átonos nalgúns complexos verbais do galego. *Estudos de Lingüística Galega*, Volume especial I, 61-79. <https://doi.org/10.15304/elg.ve1.3686>
- SOBRINHO PÉREZ, M. dos A. (2001). Análise comparativa da posición dos clíticos en galego e asturiano. En X. L. Regueira e A. Veiga (Eds.), *Da gramática ó dicionario. Estudos de lingüística galega* (pp. 265-277). Universidade de Santiago de Compostela, Anexo 49 de *Verba*.
- XOVE, X. (1991). Sobre a posición do pronome persoal átono en galego: eu o sei e eu seino. En D. Kremer (Ed.), *Actes du XVIIIe Congrès International de Linguistique et de Philologie Romanes. Université de Trèves (Trier), 1986* (pp. 515-529, vol. 2). Max Niemeyer.

Sempre concordamos co que pensamos. Unha aproximación á análise da concordancia baseada nos erros

FRANCISCO CIDRÁS

Instituto da Lingua Galega-Universidade de Santiago de Compostela

1. INTRODUCCIÓN

A presente contribución non pretende ser un estudo teórico sobre a concordancia e os seus problemas (asunto sobre o cal existe polo demais unha moi ampla literatura científica ao noso dispor¹), senón unha achega, cunha orientación didáctica, á práctica da análise da concordancia na aula baseada nos erros. Trátase esencialmente de propoñer unha abordaxe do tratamento que se debe dar no ensino a este tipo de escollos de concordancia problemáticas como ferramenta positiva (isto é, non centrada no erro concibido como falta) coa cal poderemos levar a cabo unha análise sintáctica consistente naquilo que esta debe ser de seu: unha reflexión significativa acerca de como funciona o noso sistema lingüístico². Esa reflexión consistirá esencialmente en tentar establecer os motivos polos cales escollemos unha determinada forma para expresar un determinado contido nunha determinada situación comunicativa e as consecuencias desa escolla fronte a outras opcións alternativas, que decidimos non usar por motivos que tamén deben ser indagados. En

¹ *Vid.* Tiberius et al. (2002), bibliografía especializada, con resumos e comentada, aínda que algo desactualizada polo paso do tempo; Corbett (2006) como monografía de referencia; Barlow e Ferguson (1988) como colectánea de artigos, xa clásica pero aínda de referencia pola relevancia das súas contribucións.

² A abordaxe proposta abéirase aos postulados da pedagogía do erro, que defende unha achega á aprendizaxe baseada no erro coa cal se procura unha adquisición do coñecemento e das competencias máis eficaz, construtiva e duradeira (*vid.* e.g. Astolfi, 1997; De la Torre, 2004).

definitiva, con esta dinámica tentaremos artellar unha análise dos erros e das escollos problemáticas que favoreza a construción do propio saber gramatical do falante (do alumnado, neste caso) de xeito que este sexa quen de transferir esa reflexión á mellora das súas propias producións e da súa competencia comprensiva. Iso é o único que dá sentido á análise sintáctica; pero dálle moito sentido.

Vaia por diante a explicitación dalgunhas notas vindicativas implícitas no texto. En primeiro lugar, como é doado de inferir, centrámonos na análise da concordancia, aínda sen ser ela unha escolla feita ao puro azar, non é mais que unha estratexia coa que reivindicamos, de maneira metonímica, a utilidade da análise sintáctica no seu conxunto como unha actividade didáctica que *ten sentido* (en todos os sentidos da palabra). Por outra banda, centrámonos nos erros de concordancia, máis alá da súa gran frecuencia e da necesidade de non esquecer o papel do ensino na construción dunha lingua de calidade, significa unha reivindicación do erro (ou, máis asepticamente, da escolla inesperada) como evidencia, é dicir, como fonte para coñecermos o sistema lingüístico, para sabermos como funciona a nosa competencia comunicativa. Finalmente, subxacendo a todas as argumentacións que aquí se sosteñen latexa aínda outra reivindicación: a dunha concepción comunicativa, funcional e de base cognitiva da gramática como (a) unha entidade que emerxe do uso e que, por tanto, só pode ser estudada no uso, e asemade como (b) un sistema semántico, orientado á interacción comunicativa, que vai acompañado dun sistema formal só porque non cabe outra opción. En efecto, as formas lingüísticas son puramente instrumentais con relación aos contidos que transmiten, pero son un instrumento indispensable para que esa transmisión teña lugar con éxito. Ás veces esquécese esa preeminencia dos contidos sobre as formas na gramática, chegando a sosterse, por exemplo, que a análise sintáctica é algo radicalmente distinto a unha análise semántica. De aí a soster que a análise sintáctica non ten sentido, que non serve para nada, hai un treito moi curto: a carreiriña dunha profecía autocumprida («Claro que *esa* análise sintáctica *así* non ten sentido: acabas de quitarllo!»).

Tomarase como apoio empírico os materiais do *Corpus de textos galegos escritos por estudantes no ámbito académico (CORTEGAL)*, ferramenta de análise da competencia escrita desenvolvida no Instituto da Lingua Galega e presentada en Álvarez de la Granja e Rodríguez Rodríguez neste libro. Abofé que a propia actividade docente fornece acotío exemplos abondos cos que manexarse na aula, coa vantaxe nada desprezable da súa oportunidade. Mais traballarmos cun corpus anotado ofrece o proveito alternativo de podermos

facen escollas orientadas a un tratamento do tema máis sistemático e pautado, sen renunciarmos á vantaxe da autenticidade e á utilidade do permanente acceso aos textos, que teñen que ser sempre o *locus* da análise. Ademais, o feito de tratarse de producións escritas en galego nun ámbito académico como son as probas de acceso á universidade, etapa final do ensino secundario, outórgalle á ferramenta un valor engadido e unha garantía na devandita procura de autenticidade e de identificación dos axentes indagadores co obxecto indagado. Como se sinala noutro lugar deste mesmo volume, un dos obxectivos explícitos de *CORTEGAL* é que se converta «nunha ferramenta didáctica de aplicación directa nas aulas, de tal xeito que o alumnado traballe con textos escritos por outros/outras estudantes, moi en especial no marco da aprendizaxe guiada por datos («data driven learning»)» (*vid.* Álvarez de la Granja e Rodríguez Rodríguez, §1 neste volume). A aprendizaxe guiada por datos ten unha longa traxectoria, aínda que centrada sobre todo no ensino de segundas linguas, maiormente o inglés (*vid.* e.g. McEnery e Richard, 2010; Römer, 2011); mais ten tamén numerosas aplicacións fóra dese ámbito, como se pode verificar e.g. en Gilquin e Granger (2022) ou Vázquez Rozas e Blanco (2022), que constitúe para nós unha referencia próxima onde atoparmos «un panorama actualizado de la lingüística de corpus aplicada a la enseñanza del español como primera y segunda lengua, atendiendo tanto a la vertiente investigadora como pedagógica» (p. 342). En galego é aínda un camiño pouco transitado e ferramentas como *CORTEGAL* tentan precisamente facilitar a súa exploración.

2. A IMPORTANCIA DA ANÁLISE SINTÁCTICA

Parece que corren malos tempos para a sintaxe e para a análise sintáctica na aula. Cando menos esa é a impresión que proxectan numerosas evidencias perceptibles no ambiente (incluíndo moi significativamente os ambientes docentes) que lles atribúen unha suposta e non sorprendente inutilidade e unha notable responsabilidade nos inadecuados enfoques dominantes no ensino de linguas. A práctica habitual da análise sintáctica na aula non é, con certeza, allea a esa responsabilidade, mais o xuízo ten tamén o seu punto de tópico acomodaticio que cómpre analizar e matizar. O asunto salta con frecuencia do ámbito académico ou docente mesmo ás páxinas dos xornais e á controversia da rúa. Como mostra citarei un relativamente recente titular do diario *El País* (08/11/2021) que, coa peza xornalística enteira, comento nas miñas aulas de introdución á gramática na universidade. Di así:

Menos análisis sintáctico y más aprender a comunicarse

El Gobierno plantea un vuelco a la enseñanza de la Lengua

O titular capta contundentemente a atención do lector ou lectora (é algo que se pide a todo titular!) e non semella o mellor reclamo para comezar unha introdución á sintaxe. Mais as cousas non son case nunca o que parecen. Quen completa a lectura atenta da peza decátase diso axiña ao ler as declaracións das dúas expertas responsables da proposta, veteranas profesoras de instituto, que constitúen a fonte informativa da noticia: contra a presentación inicial do proxecto como «un cambio profundo en la manera en que se enseña la asignatura», as expertas consultadas «subrayan que el planteamiento no es en realidad tan novedoso», o que se aclara finalmente coas explicacións literais que se reproducen delas:

«Venimos de una tradición que ha reservado al análisis sintáctico un lugar hegemónico dominado por rutinas de clasificación, etiquetaje y procedimiento. Horas y más horas de práctica en las que el análisis y sus convenciones se convierten en un fin en sí mismo», dice Linares.

«La construcción de un saber gramatical es esencial», sigue Jover, «pero ha de suscitarse a través de la reflexión de los estudiantes: la formulación de hipótesis, la búsqueda de contraejemplos, las generalizaciones... Queremos que sean capaces de transferir esos aprendizajes a la mejora de sus propias producciones, a la solución de problemas de comprensión, a la mejor adecuación entre lo que quieren decir y lo que efectivamente dicen».

As declaracións resultan irreprochables e sitúanse claramente na liña da concepción da análise sintáctica que acabo de soster no comezo desta achega. Lonxe de revelaren unhas *syntax haters*, amosan unha evidente afinidade coas posicións doutras persoas que defenden a importancia da análise sintáctica e do saber gramatical no ensino da lingua ao tempo que se laian da deturpación que esa análise amosa moi maioritariamente na súa posta en práctica. As declaracións que acabo de reproducir mostran unha máis que posible filiación cunhas coñecidas reflexións que sobre o tema facían hai máis de 30 anos M.^a Lluïsa Hernanz e José M.^a Brucart nun manual universitario —*La sintaxis* (1987)— que gozou, e aínda goza, de gran difusión e prestixio. Velaí o teor deste texto, que tamén analizo nas miñas aulas de sintaxe e agora produzo para deixar patente esa comunidade de reflexións:

Uno de nuestros objetivos principales es el de reivindicar un concepto amplio de análisis sintáctico, en el que el criterio decisivo sea el de la argumentación. Por desgracia, a veces se confunden en este ámbito los medios con los fines, hasta el

punto de identificar los esquemas gráficos de representación (diagramas arbóreos, paréntesis etiquetados, etc.) con el análisis propiamente dicho. Tal error lleva en la práctica a concebir esta actividad como un ejercicio puramente mimético desprovisto de todo interés, plagado de decisiones arbitrarias y en cuya realización el estudiante renuncia de antemano a utilizar su intuición de hablante. Creemos, por el contrario, que el análisis debe concebirse primordialmente como un ejercicio de reflexión intelectual sobre el funcionamiento del sistema lingüístico. Para ello, el estudiante ha de contar con dos armas fundamentales: la intuición, derivada de la competencia que tiene como hablante de la lengua, y la formación teórica. Adquirir esta última equivale, en último término, a aprender los rudimentos de la argumentación lingüística. De ahí la importancia pedagógica que otorgamos a esta actividad, pues es la que en mayor medida puede evitar que los árboles no nos dejen ver el bosque de la sintaxis (Hernanz e Brucart, 1987, p. 8).

En realidade, as orixes do discurso antigramatical remóntanse aos anos 60 do século pasado, como reacción á centralidade que daquela estaban a adquirir os modelos teóricos formalistas. Non é por tanto «tan novedoso» ese que na altura si foi un «cambio profundo en la manera en que se enseña la asignatura», propiciado por defensores dos modelos de ‘linguaxe en uso’. Entre os seus pioneiros destaca como referente a figura de M.A.K. Halliday, quen no seguinte texto lembra ese camiño ao tempo que lamenta con certo amargor as oscilacións pendulares, os movementos de refluxo e o eco que este tipo de dinámicas adoita xerar:

Whatever the ultimate goal that is envisaged, the actual analysis of a text in grammatical terms is only the first. (...) But whatever the final purpose or direction of the analysis, there has to be a grammar at the base. Twenty years ago, when the mainstream of linguistics was in what has been called its ‘syntactic age’, it was necessary to argue against grammar, pointing out that it was not the beginning and the end of all study of language. (...) Now however it is necessary to argue the opposite case, and to insist on the importance of grammar in linguistic analysis. If now I appear as a champion of grammar, it is not because I have changed my mind on the issue, but because the issue has changed. The current preoccupation is with discourse analysis, or ‘text linguistics’; and it is sometimes assumed that this can be carried on without grammar — or even that it is somehow an alternative to grammar. But this is an illusion. A discourse analysis that is not based on grammar is not an analysis at all (Halliday, 1985, pp. XVI-XVII).

Os modelos formalistas e gramaticocéntricos que tiveron unha enorme difusión hai sesenta anos foron daquela trasladados axiña, de maneira non sempre atinada, ao ámbito escolar. Unha das súas herdanzas máis marcantes foi a xeneralización do uso dos diagramas arbóreos, que chegaron a identificarse, malfadadamente, coa propia análise sintáctica. Non serei eu quen rexeite nin quen enxalce este modelo de representación gráfica das análises. Posúe

claras vantaxes, tanto de visualización como de protocolarización e pautado do procedemento de análise. Pero é o certo que tanto protocolo e tanta pauta acabaron por afogar a esencia da análise: a reflexión sobre os propios usos lingüísticos, baseada na intuición e na argumentación, que foi substituída por un exercicio non-significativo consistente en restras de decisións que para a maioría das persoas son arbitrarias, carecen de sentido e constitúen un fin en si mesmo, como se sinalou máis arriba. A centralidade destas prácticas de análise-sen-análise é, desde hai ben tempo, unha rareza nas aulas, con contadas excepcións. As prácticas, porén, persisten de maneira bastante xeneralizada, ben que sen dedicarlles xeralmente «horas e máis horas» na orde de traballos cotiáns da aula.

Nos meus longos anos, hoxe pasados, como responsable das PAU/ABAU na materia de Lingua Galega e Literatura (nas que o peso das preguntas gramaticais é propositadamente pequeno) puiden constatar un inquietante paradoxo: de formular preguntas co enfoque clásico da análise sintáctica, a reacción máis xeral, común a alumnado e profesorado, era do tipo «sei o que che teño que responder, pero non entendo que sentido ten que me fagas esa pregunta»; de formular, alternativamente, propostas de análise cun enfoque comunicativo, con máis reflexión menos pautada, a reacción era do tipo «entendo o sentido do que me preguntas, pero non sei ben que é o que queres que che responda». O panorama coa análise sintáctica é, desde esta miña percepción, desacougante: ou «sei como se fai pero non ten sentido» ou «ten sentido pero non sei como se fai»... Obviamente iso ten que ver coas prácticas habituais nas aulas, en que moitas veces facemos o que teoricamente criticamos por carente de sentido, porque descoñecemos (ou non temos suficientemente pautadas) estratexias alternativas. Claro que o exercicio de pautas nos coloca de novo no camiño dos riscos que acabamos de lamentar...

3. A IMPORTANCIA DA ANÁLISE DOS ERROS

Nesa tarefa de reflexión argumentativa e intuitiva acerca dos usos lingüísticos en que consiste a análise sintáctica, o papel dos erros ten unha grande importancia. Na investigación son tan relevantes como os acertos, pois uns e outros axudan igualmente a coñecer como funciona o sistema de escollas que é a gramática. No traballo da aula, particularmente no ensino secundario, son seguramente máis importantes, por un dobre motivo: dunha banda pola súa relevancia como evidencia na construción do coñecemento lingüístico; pero ademais porque ese coñecemento é fundamental para o control dos propios erros e, por tanto, para a consecución da lingua de calidade a que a escola debe contribuír.

É obvio que o concepto de erro de lingua (ou de calquera outro tipo) está impregnado de connotacións sociais negativas ao vincularse á idea dunha falta, dunha equivocación, de facer algo que non se debería facer. Cómpre manter distancia con este tipo de connotacións, de natureza ideolóxica, non factual³. Concibiremos daquela o erro como unha *escolla inesperada*, o que nos coloca nun terreo cientificamente máis neutro. Incurrir nun erro non é daquela facer algo que non se debería facer, senón facer unha escolla distinta da esperada pola maioría da comunidade de falantes. É escusado dicir que, na fin de contas, arredarse do comunmente esperado adoita non pasar desapercibido e acostuma ter algún tipo de consecuencia (case sempre problemática) na comunidade.

Esta referencia ao «comunmente esperado» incide tamén no carácter relativo do que se entende por erro, pois a gramática (i.e., o código prescriptivo de escollas) non ten unha densidade homoxénea en toda a súa topografía. John Du Bois (1985, p. 363) resumíao dunha maneira tan breve como acertada: «Grammars code best what speakers do most». O da concordancia é, con certeza, un eido privilexiado en que constatar como a presenza de variabilidade de escollas en competencia ás veces está asociada ao concepto de erro de maneira moi contundente, outras de maneira máis tenue e en ocasións está completamente dissociada del, ao non existir un «ditame social» minimamente consolidado ao respecto. Sobre este tema volveremos máis adiante.

No alicerce da importancia que estamos a atribuír ás escollas erradas na análise e indagación sintácticas encóntrase unha hipótese moi estendida que o neurocientífico cognitivo John Morton formulou co seguinte principio: «Errors do not just happen, but are caused» (Morton, 1964, p. 41). Para alén de que os principios xerais son sempre susceptibles de seren matizados nos detalles miúdos, o certo é que a indagación desas causas dos erros de lingua nos coloca no centro da investigación lingüística. Neste ámbito científico concreto existe de vello unha nutrida corrente que traballa desde a idea de que os erros son unha fonte caudalosa de evidencias para coñecermos como funcionan as linguas: *Speech errors as linguistic evidence*, por remedar o título dunha coñecida colectánea de contribucións clásicas na literatura científica sobre o tema (Fromkin, 1984).

³ Porén, ao mesmo tempo convén non esquecer que o propio concepto de *lingua de calidade* (depurada e desprovista de erros), como o de *corrección lingüística*, o de *norma* ou o de *estándar*, son pola súa natureza construcións ideolóxicas. Non podendo ser asépticas, cabe só esixir que sexan lexítimas. A gramática é a institución social encargada de fixar e lexitimar eses canons, e a escola a de espallalos e perpetualos, dentro dun cadro xeral en que se asume que educar consiste en inducir a cumprir as normas.

A elaboración de repertorios de erros ou anomalías no uso da lingua foi, desde antigo, unha actividade moi concorrida, porque, como acabamos de dicir, as escollas inesperadas (non só as lingüísticas, pero tamén elas) son sempre obxecto de particular atención e escrutinio, unhas veces como obxecto de burla, outras de censura, outras con vontade correctiva ou didáctica. Eses repertorios son unha fonte documental de gran valor para o coñecemento lingüístico. Pensemos por exemplo no *Appendix Probi*, a coñecida listaxe de 227 erros recadada na primeira metade do século IV, xuntamente coa correspondente forma correcta (a xuízo do autor, non sempre certo), que nos bota luz sobre a evolución do latín vernáculo daquel momento cara ao protorromance. Por iso unha disciplina en que esta corrente tivo desde sempre moito vigor é a gramática histórica (Sturtevant, 1917; Jespersen, 1922). A Sturtevant débemoslle a definición dos erros de lingua como «an unintentional linguistic innovation» (1947, p. 38), considerándoos por tanto non só un rexistro senón un motor do cambio lingüístico.

Outra disciplina en que se manifesta de maneira moi patente a importancia do estudo dos erros de lingua é a fonética, o que parece vincularse co feito de que a maioría dos repertorios clásicos de lapsus línguae (desde o citado *Appendix Probi* aos de Meringer e Meyer e aínda outros máis modernos) centraban a súa atención maiormente en erros de pronuncia. Os traballos senlleiros de Victoria Fromkin e os alleos incluídos na súa citada recompilación (Fromkin, 1984) son bo exemplo desta tendencia, mais ao mesmo tempo mostra de que os erros de lingua fornecen evidencias para o coñecemento de todas as disciplinas lingüísticas e son unha ferramenta moi importante na construción dunha teoría xeral da lingua, pezas chave para completar o quebracabezas da linguaxe, como se pode verificar en Fromkin et al. (2011). Tamén no *Fromkin's Speech Error Database*, iniciado pola autora hai medio século e aínda en desenvolvemento, onde se recollen e clasifican erros de lingua en todos os niveis, incluíndo o sintáctico⁴. Por outra banda, todos os cörpera modernos con anotación de erros inclúen e clasifican rexistros pertencentes ás distintas compoñentes (excepción feita da textual ou discursiva, que non adoita estar anotada), con atención máis ou menos específica aos erros de sintaxe. Tamén o fai así *CORTEGAL*, que clasifica os erros de lingua nas seguintes categorías⁵:

⁴ *Fromkins Speech Error Database — Max Planck Institute for Psycholinguistics*. https://www.mpi.nl/dbmpi/sedb/sperco_form4.pl [última consulta: 10/09/2022].

⁵ Cabe observar que a maioría dos cörpera con anotación de erros son recompilacións de textos de segundas linguas ou linguas estranxeiras, e están orientados ao estudo dos procesos de adquisición destas (*CORTEGAL* é neste senso unha excepción). Por contra, as colectáneas e bases de datos adoitan

- 1) Problemas ortográficos
- 2) Problemas morfolóxicos
- 3) Problemas léxicos
- 4) Problemas gramaticais [= sintácticos]
- 5) Problemas semántico-pragmáticos
- 6) Problemas discursivos

A categoría «problemas gramaticais» de *CORTEGAL* recolle en realidade unicamente problemas sintácticos (i.e., problemas relativos á construción de frases e oracións). Todos eles van convenientemente etiquetados con candaseu código. No seu estado actual, o corpus identifica un total de 23.169 etiquetas de formas non estándar, das cales 3958 (17,08%) corresponden a ocorrencias de problemas de sintaxe. Entre elas ocupan un lugar destacado os erros de concordancia.

4. A IMPORTANCIA DA CONCORDANCIA... E DOS SEUS ERROS

A concordancia é un mecanismo enormemente importante na gramática da maior parte das linguas, en particular das flexivas. Tamén é «un dos paraísos da variabilidade lingüística» (Cidrás, 2009, p. 33), o que propicia a ocorrencia, no seu manexo, de numerosas escollas erradas ou problemáticas, alén de moitas situacións de escolla múltipla en que non existe un «ditame social» consolidado que permita discriminar entre o «errado» e o «correcto». O seu manexo está regulado por unha confluencia de factores morfosintácticos, semánticos e pragmáticos, nunhas interrelacións que son en realidade comúns ao funcionamento global da gramática, mais que aquí se tornan particularmente obvias. Por todo iso é unha zona da gramática de grande interese para a reflexión e análise sintácticas en xeral, pero máis aínda desde a «abordaxe-centrada-no-erro» que aquí adoptamos⁶.

referirse a erros no manexo de primeiras linguas (como é o caso da de Fromkin que se acaba de referir). Por suposto, dunhas e das outras se poden extraer conclusións de interese, pero é obvio tamén que existen diferenzas relevantes sobre o tipo de erros predominantes en cada caso.

⁶ Porén, non existe no ámbito da lingüística nin na gramaticografía galegas unha reflexión forte sobre a importancia da concordancia ou o tratamento dos seus erros de uso. A ausencia explícase con certeza polo feito de estes asuntos non formaren parte do reducido grupo de trazos identitarios que singularizan o galego, na cerna diso que se denomina «lingua de calidade» (Freixeiro Mato, 2009), identificada tactivamente cun modelo de lingua depurado das interferencias asimilacionistas do español, que non teñen na concordancia un obxectivo urxente. Situándonos supletoriamente no ámbito máis próximo e máis normalizado, Raposo et al. (2021) é un bo exemplo da importancia recoñecida á concordancia como mecanismo na gramaticografía portuguesa máis recente, e Peres e Mória (1995) do seu enorme relevo nas «áreas críticas» da lingua, aquelas onde medran a variación, a diversidade de opcións e por tanto os problemas de escolla, cuxo bo manexo contribúe tamén —cómpre non esquecerlo— á calidade da lingua.

A sintaxe ten o seu principal foco de interese na análise das relacións gramaticais, i.e., dos tipos de vínculos que establecen entre si as palabras e as combinacións de palabras para producir enunciados. A peza chave nesa análise son as marcas desas relacións gramaticais, que responden basicamente a só tres grandes tipos: concordancia, marca de caso e orde de palabras. Diferéncianse polo lugar que ocupa a marca en cada un deles e responden a cadansúa estratexia distinta atendendo á súa función primaria, aquela para a que foron de seu concibidas (a polivalencia funcional é unha característica recorrente nos sistemas lingüísticos; *vid.* Cidrás, 1992 para máis detalles sobre eses tipos de marcas en galego; cf. Croft, 1988):

- a) *Concordancia*: a marca vai localizada *ex situ* (i.e., fóra do seu lugar natural), acaroada ao núcleo predicativo —verbal ou nominal— de que dependen os constituíntes dos cales se marca a relación; a concordancia verbal responde de seu a unha estratexia cohesiva, de recuperación referencial no nivel textual.
- b) *Marca de caso*: a marca vai localizada *in situ*, acaroada ao nominal do cal se marca a relación; responde a unha estratexia propiamente relacional, de identificación de relacións sintácticas no nivel oracional.
- c) *Orde de palabras*: é unha marca *sine situ*, xa que carece de matriz morfolóxica; outórgaselle á disposición secuencial dos constituíntes valor de marca das relacións que manteñen; responde de seu a unha estratexia informativa, de identificación de funcións pragmáticas no enunciado.

A *orde de palabras* ten en galego un papel moi marxinal e subsidiario como marcador de funcións sintácticas, ao punto de que o seu uso na práctica non dá lugar a erros. Un exemplo potencial de escolla de orde de constituíntes supostamente inaceptable sería, e.g., unha construción do tipo *O neno viu a nena* en que *o neno* é complemento directo e *a nena* suxeito (i.e., en que é a nena quen ve o neno), fronte á máis canónica construción homógrafa en que *o neno* é suxeito e *a nena* complemento (i.e., en que é o neno quen ve a nena). Porén, a primeira construción, aínda que infrecuente, é tamén correcta e posible, sobre todo na lingua oral, e non hai ambigüidade entre ambas, porque unha e outra se diferencian mediante marcas prosódicas; a primeira presenta o obxecto focalizado en posición inicial (*focus fronting*); a segunda, tema inicial, que é a opción non marcada. Non hai, pois, erro na orde de palabras.

A *marca de caso* é, dos tres mecanismos en causa, o único que está deseñado especificamente, e non de maneira secundaria, para o marcado de relacións gramaticais, como evidencia o feito de aparecer a marca acaroada ao

constituínte cuxa función se quere marcar. Escúsase lembrar que o galego perdeu a marca de caso flexiva e que só posúe a non-flexiva ou adposicional, nomeadamente as preposicións. O deseño específico do mecanismo fai que resulte nun marcado inequívoco e tamén nun número relativamente baixo de erros de uso (fronte á concordancia, en ambos os casos). De feito, a maioría dos usos inaceptables deste mecanismo se deben a factores esóxenos, de interferencia doutros sistemas lingüísticos, non a problemas derivados das características do sistema de marcado. Tal ocorre co que é, con moita diferenza, o tipo máis frecuente de erro correspondente a este mecanismo: o uso indebido de preposición co complemento directo, case sempre por abuso da marca *a*, producido por interferencia do español. Por exemplo⁷:

- (1) Os restaurantes contratan **a** moita xente sen estudos
[en lugar de Os restaurantes contratan moita xente sen estudos]
- (2) Moitas empresas están pechando e botando **aos** seus traballadores
[en lugar de Moitas empresas están pechando e botando os seus traballadores]
- (3) Deberíamos [...] darlle unha oportunidade a pequenas empresas, que seguramente non exploten **á** xente pobre
[en lugar de deberíamos [...] darlle unha oportunidade a pequenas empresas, que seguramente non exploten a xente pobre]
- (4) O mesmo pasa cos rapaces ao ver **ao** seu correspondente equipo xogar pola televisión
[en lugar de O mesmo pasa cos rapaces ao ver o seu correspondente equipo xogar pola televisión]

A *concordancia* é un mecanismo deseñado para recuperar entidades presentadas noutro lugar do texto sen necesidade de reiteralas lexicamente cada vez que estas volven aparecer no discurso⁸. Ela é a que permite, por exemplo, a través da rica flexión verbal de número e persoa, que o galego sexa unha lingua de suxeito nulo, i.e., que este constituínte acostume aparecer elíptico cando se trata dunha entidade que xa fora introducida no discurso previo ou é recoñecible na situación comunicativa, o cal permite alixeirar o discurso e facilita o seu procesamento. A concordancia concíbese por tanto como un recurso cohesivo, no nivel textual, aínda que secundariamente se recicle, no oracional, como marca da relación gramatical que os participantes na predición establecen co seu predicador. Esa función cohesiva primixenia explica

⁷ Todos os exemplos numerados que aparecen neste traballo están tirados do corpus *CORTEGAL*, agás que se especifique o contrario. Ademais, introdúcense ao longo do texto algúns outros creados *ad hoc* polo autor a efectos de claridade expositiva.

⁸ A concordancia verbal ou textual. A nominal ou frásica é, como a marca de caso, un mecanismo concibido cunha estratexia relacional, de marcado de relacións gramaticais, pero só no dominio da frase. *Vid.* §5.

por que a marca de concordancia se aloxa no verbo, vista a condición que este ten de predicador que selecciona os argumentos da predicación e lles asigna cadanseu papel semántico. Dado que ese papel debe ser identificado dalgún xeito, aprovéitase a propia marca cohesiva de concordancia habilitándoa como marca de relación gramatical ou, se se prefire, como marca de función sintáctica, coa cal se identifica indirectamente o papel semántico dos argumentos (Cidrás, 2012).

A localización no verbo da marca de concordancia era, como marca cohesiva no dominio textual, a opción natural; porén, na súa dimensión habilitada como marca de relación no nivel oracional, trátase dun marcado *ex situ*, ao estar aloxado fóra do seu *locus* natural (i.e, a marca, a carón do constituínte que mantén a relación marcada). A mecánica de funcionamento da concordancia, á que nos referiremos máis adiante, e a «deslocalización» da marca no verbo, fóra do seu *locus* natural a carón do constituínte que mantén a relación, outórganlle á concordancia un sensible grao de incerteza no mercado da relación e unha notable variabilidade nas opcións de concordancia, que concorren en competencia. Estes factores acaban por propiciar a aparición de escollas problemáticas, moitas delas consideradas erros. A concordancia é, en efecto, o mecanismo de marcaxe de relacións gramaticais cun maior risco potencial de erro ou de conflito.

Os datos cuantitativos de *CORTEGAL* aboan empiricamente o estado de cousas que, nun plano teórico, acabamos de describir:

— Ocorrencias totais de problemas de sintaxe:	4017	(100,00%)
— Problemas de marcado de relacións gramaticais:	1802	(44,86%)
dos cales:		
- Problemas de concordancia:	1159	(64,32%)
- Problemas de marca de caso:	643	(35,68%) ⁹
- Problemas de orde de palabras:	0	(0,00%)

Obsérvase, en primeiro lugar, a centralidade dos problemas de marcado das relacións gramaticais no conxunto dos problemas de natureza sintáctica (gramaticais na terminoloxía de *CORTEGAL*): case a metade das ocorrencias de problemas ou erros de sintaxe etiquetados corresponden ao marcado de

⁹ En *CORTEGAL* etiquétanse especificamente os problemas de concordancia, pero non os de marca de caso, que aparecen subsumidos no conxunto dos de «adición/omisión de preposición», que acadan as 697 ocorrencias. Delas cómpre detraer manualmente os usos indebidos de preposición en construcións gramaticalizadas en que esta xa non funciona como marca de caso (e.g., as preposicións usadas en perífrases verbais, en particular o uso de [*ir* (*a) + inf.], que achega a maior parte dos usos detraídos). Obtense así á cifra de 643 usos errados de marca de caso.

relacións gramaticais. O marcado destas relacións resólvese en galego, como se indicou, nun uso complementario da concordancia e a marca de caso, cun papel moi marxinal da orde de constituíntes. O estudo cuantitativo de erros etiquetados en *CORTEGAL* deseña, consonte o esperable, un claro predominio dos problemas de concordancia, que achegan case dous terzos do total, fronte a algo máis dun terzo dos problemas de marca de caso e unha nula contribución da orde de constituíntes. Cómpre advertir aínda que ese predominio está en realidade distorcido á baixa pola interferencia do español no marcado preposicional de complemento directo en galego, que acumula 282 usos errados, preto da metade do total de ocorrencias problemáticas de marca de caso. De non existir ese factor externo, alleo por completo ás debilidades e fortalezas inherentes aos mecanismos de marcaxe, o peso específico da concordancia no conxunto dos usos problemáticos sería aínda sensiblemente máis elevado:

— Problemas de concordancia:	1159	(76,25%)
— Problemas de marca de caso (sen *a+CD):	361	(23,75%)

Esa maior «problematicidade» potencial da concordancia fronte á marca de caso ten que ver co deseño e o modo de funcionar de cada mecanismo: específico e localizado no caso da marca de caso; inespecífico/habilitado e deslocalizado no caso da concordancia. É doado comprender por que o carácter non específico e desprazado da concordancia na marcaxe de relacións gramaticais lle outorga maior risco potencial de escollos erradas e problemáticas¹⁰: fronte a un marcado de caso realizado *in situ*, concordar consiste en escoller as marcas morfolóxicas variables dunha palabra en función das doutra situada noutro lugar da cadea lingüística que pode mesmo estar moi distante. Este exercicio de covariación está, ademais, deseñado especificamente como un

¹⁰ Repárese en que ese carácter máis potencialmente problemático da concordancia (por motivos ademais estruturais, de deseño do mecanismo) non se traduce nun menor uso fronte á marca de caso, senón todo o contrario, como evidencian sincronicamente o seu maior uso cuantitativo e diacronicamente o avance na gramaticalización e uso dos clíticos como morfemas de concordancia fronte á perda da marca de caso flexiva. Este aparente paradoxo explicase porque a concordancia presenta asemade importantes vantaxes de polivalencia funcional que compensan sobexamente aquela eiva. Trátase de vantaxes compartidas nos niveis da sintaxe oracional e da cohesión textual, en que a marca de caso é irrelevante. Ademais, a distribución complementaria no uso da concordancia e da marca de caso vincula a primeira, *grosso modo*, co marcado relacional dos argumentos e a segunda co dos adxuntos. A diversidade de papeis que cómpre identificar nos adxuntos é de certo maior que nos argumentos pero, inversamente, estes son moito máis frecuentes que aqueles, por seren máis centrais e recorrentes na conceptualización das mensaxes e na construción de significados. Todo o cal redundna na maior importancia cuantitativa da concordancia como mecanismo de codificación gramatical en linguas como o galego.

recurso de cohesión do texto, plano no cal a identidade formal das categorías concordantes é moi pouco relevante. O resultado é un procedemento moi útil pola súa polivalencia funcional, mais que permite un alto grao de incerteza e falta de seguridade na adopción de escollas. Ora ben, establecido isto con carácter xeral, cómpre advertir que non todas as ocorrencias de concordancia son iguais en canto ao grao de especificidade de deseño nin de distancia potencial entre os termos covariantes. Consecuentemente, tamén non todas inducen o mesmo grao de incerteza nas escollas nin presentan o mesmo nivel de erros.

5. DOUS TIPOS DE CONCORDANCIA (E A AGRAMATICALIDADE DOS ERROS DUN DELES)

5.1. Diferéncianse dous tipos de concordancia ben distintos segundo o dominio en que actúan, a estratexia a que responden e os constituíntes que un e outro (ou unha e outra?) seleccionan como controlador e *target*¹¹:

- a) A *concordancia verbal* responde, como xa se sinalou, a unha estratexia cohesiva, despregada no dominio do texto, mais asemade está habilitada secundariamente como marcadora da función sintáctica dos argumentos no dominio da oración. Neste caso os argumentos son os controladores da concordancia e o núcleo verbal é o seu *target*. Por exemplo (en 5 e 6, exemplos do autor, controlador en versais, *targets* subliñados, morfemas concordantes en grosas):

(5) —Ti viches *AS NENAS*? —Eu *atopeinas* onte. *Andaban* pola rúa¹²

- b) A *concordancia nominal* responde, por contra, a unha estratexia especificamente relacional, pero limitada ao dominio da frase: identifica, dentro dela, os elementos que determinan ou complementan o seu núcleo; a maneira como o fai permítelle actuar secundariamente como un delimitador da extensión da frase, marcando o alcance da proxección do núcleo da frase (i.e., ata onde chega o seu goberno). É completamente irrelevante no dominio

¹¹ Chamamos *dominio* ao ámbito en que se dá a concordancia, o marco en que se resolven os vínculos entre os elementos concordantes (e.g., a frase, a oración, o texto); o *controlador* é o elemento que determina a concordancia; o *obxectivo* ou *target* (está moi espallado o uso termolóxico do anglicismo) é o elemento determinado, aquel cuxa forma muda para concordar cos trazos do controlador (Corbett, 2006).

¹² Obsérvese que analizamos os clíticos como morfemas de concordancia verbal a pesar de non formaren parte da flexión do verbo (análise análoga á das preposicións como marca de caso tamén nonflexiva, no ámbito nominal). A consideración dos pronomes átonos como morfemas verbais de concordancia de obxecto nas linguas romances, complementaria da flexiva de suxeito, é unha idea sostida de longa data (e.g. Lenz, 1920) e hoxe amplamente espallada (e.g. Bossong, 1980; Brown e Sempere, 1987; García-Miguel, 1991; Cidrás, 1992; Franco, 2000).

Tomemos a definición de concordancia que ofrece Susan Steele, espallada e dalgún xeito canonizada por Corbett (2006, p. 4): «The term ‘agreement’ commonly refers to some systematic covariance between a *semantic or formal* property of one element and a *formal* property of another» (Steele, 1978, p. 610; cursivas nosas). Á vista dos exemplos anteriores, cando menos en linguas como o galego cabería especificar a devandita aserción axustándoa a cada tipo de concordancia:

- a) A *concordancia verbal* consiste nunha covariación sistemática potencialmente lábil entre unha propiedade semántica ou formal do elemento controlador e unha propiedade formal correspondente do *target*.
- b) A *concordancia nominal* consiste nunha covariación sistemática fixa entre unha propiedade formal do elemento controlador e a propiedade formal correspondente do *target*.

A falta de variabilidade potencial, en condicións normais de produción, da concordancia nominal explícase polas propiedades funcionais e operativas deste tipo de concordancia, que contrastan fortemente coas da concordancia verbal: i) é un mecanismo especificamente relacional (non habilitado, como é a concordancia verbal) orientado ao marcado formal da estrutura da frase, non á aceptabilidade discursiva do texto; ii) ten como estrito dominio de actuación a frase, que presenta un alto grao de codificación gramatical (vs. o texto e o enunciado, categorías de maior pegada semántica e pragmática e maior laxitude formal); e iii) a distancia entre o controlador e o *target* é case sempre mínima (de feito adoitan aparecer en estrita adxacencia), o que deixa pouco lugar para que se dilúa o peso das propiedades formais sobre as semánticas e conceptuais, en caso de desaxuste entre unhas e outras.

5.2. En *CORTEGAL* non aparecen marcadas con etiquetas distintas as ocorrencias de concordancia verbal e nominal, polo que non é posible facer procuras específicas sobre cada unha delas. Porén, unha procura xeral ou unha orientada á selección inadecuada de xénero deita resultados teoricamente inesperados como os transcritos a seguir. Non realicei o cómputo manual sobre o total de ocorrencias problemáticas etiquetadas, mais o certo é que, sendo poucas, non son casos insólitos:

- (10) Na sociedade actual a produción e o consumo teñen un lugar destacado, por ser os pilares **da** capitalismo
- (11) Isto trátase **da** principio de demanda, ti si ves que algo é moi solicitado e ten éxito, vas producir máis
- (12) Se non hai competencia polos prezos, **as** prezos tenderán a subir

- (13) Algunhas persoas [...] van ó máis barato aínda que teña peor calidade, mentres que a outras dalle igual o prezo e guíanse **polo** súa calidade
- (14) Esta epidemia gastronómica [...] débese **o** popularidade que adquiriu a comida nos últimos anos
- (15) Ademais de supoñer un pracer comer **un** boa comida, tamén e bo para a nosa saúde
- (16) Quero distinguir dúas categorías dentro da gastronomía **dun** zona
- (17) Na sociedade actual vívese **un** loita constante por ter máis ca os demais

Como explicar estas ocorrencias que acabamos de considerar agramaticais? Penso que a pesar da súa presenza non illada, incontestable e incontornable no corpus, ningunha persoa galegofalante cuestionará ese xuízo e todas concordarán en que a súa gramática non lles permite producir tales escollas en condicións de uso normais. Con seguridade, ningunha das propias persoas que escribiron eses exemplos diría **a capitalismo*, **o súa calidade*, **o popularidade*, **un boa comida*, **un loita*; por que o escriben daquela nun texto académico que saben que vai ser corrixido avaliando a súa calidade lingüística? Repárese no ex. (12) en que lemos **as prezos* inmediatamente despois de ver escrito *os prezos*. Na miña opinión, é a propia dinámica da produción de textos escritos a que favorece a aparición destes usos. Nesa dinámica son comúns rupturas e discontinuidades no proceso de construción textual, non sempre ben resoltas cando este se retoma.

É moi elocuente ao respecto a análise das emendas ou riscadas que con frecuencia acompañan, nos textos manuscritos, estas ocorrencias, e que *CORTEGAL* permite rastrexar seleccionando a opción de visualización «transcrición completa». Así, por exemplo, en (10) podemos ver que o autor ou autora do texto inicialmente escribira *os pilares da soc*, seguramente pensando na *soc[iedade]*; porén, muda sobre a marcha de idea, risca a palabra inconclusa e substitúea por *capitalismo* (escolla que razoablemente considera máis apropiada e que lle permite evitar unha pouco elegante reiteración léxica) sen mudar o artigo feminino que correspondía á opción léxica inicial. De maneira análoga, en (13) a persoa que redacta o texto escribe primeiro *pola súa*, logo corrixe para *polo seu* e finalmente (cando se decide polo termo *calidade*) volve escribir atinadamente *súa...* pero deixando atrás sen corrixir o artigo en masculino. A persoa que en (12) escribe *Se non hai competencia polos prezos...* ten probablemente planificada na súa mente unha entidade feminina plural como punto de partida da oración principal, e avanza esta cun artigo concordante con ese novo controlador electo; porén, opta finalmente por repetir o nominal *prezos*, mais non recúa para axustar a forma do *target* a este controlador da concordancia: *Se non hai competencia polos prezos, as [---] prezos tenderán a subir*. Na lingua oral tamén se dan

reformulacións do enunciado, e abofé que moitas, pero neste caso a solución pasaría por reformular toda a FN, pois o artigo é prosodicamente un clítico do nominal: *Se non hai competencia polos prezos, as... os prezos tenderán a subir*. Na escrita, o lapso reformulatorio, probablemente longo, fai perder a percepción unitaria da categoría dominio. Resultan así construcións que son agramaticais porque son produto dunha cancelación transitoria da gramática. Non existe unha motivación cognitiva que conduce a unha escolla gramaticalmente discordante ou non canónica, senón unha quebra do proceso produtivo do enunciado, que remata mal suturada.

Existen varias evidencias do carácter anómalo destes usos, non susceptibles de gramaticalización por estaren construídos á marxe de calquera gramática posible. Entre eles cómpre destacar o feito de que a discordancia, cando se dá, afecta exclusivamente aos predeterminantes da frase (e xeralmente só ao primeiro deles, como se ve nos exemplos), sen estenderse aos demais *targets* da concordancia. Non atopamos no corpus frases do tipo **un bo comida* (cf. 15), pero sobre todo sería virtualmente imposible atopar outras do tipo **un comida bo* ou **un bo comida mexicano*. Tal demostra que estamos diante dunha produción truncada e reformulada de xeito incongruente coa gramática, mais esa incongruencia só afecta ao ataque da frase, ao(s) elemento(s) xa escrito(s) antes da ruptura reformulatoria. Outro indicio que con certeza apunta na mesma liña é o feito de que este tipo de concordancias ocorren sempre (ou case sempre) por unha selección inadecuada de xénero, cando no conxunto dos erros de concordancia a causa esmagadoramente maioritaria do problema é a selección inadecuada de número (*vid. infra* §6.2). Isto é así porque as reformulacións e cambios sobrevidos na escolla de controlador, que están na orixe destas concordancias nominais feitas á marxe da gramática, afectan frecuentemente ao xénero do controlador (se muda entre o da escolla inicial e a final), pero case nunca ao número, pois é pouco factible que a reformulación afecte á cantidade de entidades reais referidas, xa que muda só a escolla léxica.

6. OS «ERROS GRAMATICAIS» DE CONCORDANCIA COMO DESAXUSTES ENTRE A CONCEPTUALIZACIÓN SEMÁNTICA E A CODIFICACIÓN GRAMATICAL

6.1. Adecuación formal ou congruencia semántica?

A inmensa maioría das ocorrencias de concordancia erradas, problemáticas ou non canónicas son produto e manifestación dalgún tipo de desaxuste entre a conceptualización semántica da mensaxe e a súa codificación gramatical na oración. Estas faltas de correspondencia (*mismatches*) entre

contido semántico e forma gramatical, ou entre forma e función, son unha peza chave para comprendermos o funcionamento non só da concordancia (como salienta Corbett, 2006 entre outros), senón do sistema lingüístico no seu conxunto (Bresnan, 1995; Francis e Michaelis, 2003; Yuasa, 2005). Lonxe de seren unha pedra no zapato, constitúen unha compoñente esencial da arquitectura da gramática e, pola ostensible incongruencia que amosan, chaman a atención das persoas e suscitan reflexión e debate nese lingüista que todos os falantes levamos dentro. Non deben daquela desaproveitarse na aula para levar a cabo, a partir delas, unha análise sintáctica merecedora do seu nome.

No caso da concordancia, o desaxuste non se produce no mecanismo en si (i.e., na correspondencia entre controlador e *target*) senón internamente no seo do controlador (e.g., *a xente*: morfoloxía singular para un contido de pluralidade de persoas) e resólvese nos seus *targets* mediante a concorrencia de varias solucións entre as que o falante debe escoller unha; e.g.: *A xente {quéixase / quéixanse / queixámonos / queixádesvos} cando {lle / lles / nos / vos} preguntan pero á hora da verdade {cala / calan / calamos / calades}*. Cada unha delas é funcional, en tanto que «resolve» o desaxuste dunha determinada maneira; pero cada maneira ten implicacións semánticas ou pragmáticas distintas. Todas as opcións son gramaticais, en tanto que son respostas sistemáticas e significativas que a gramática ofrece para resolver unha inconsistencia do sistema. En ningunha das variantes dispoñibles hai ausencia de concordancia entre o controlador e os seus *targets*. Outra cousa é o xuízo prescriptivo que cada unha suscita: algunhas escollas son mellor aceptadas que outras... e algunhas son abertamente consideradas erros. As fronteiras entre o errado e o correcto amósanse difusas e moitas veces caprichosas, mesmo contraditorias (o que non estraña, pois ao cabo son decisións sociais). Pero o mecanismo é sempre o mesmo, á marxe do xuízo prescriptivo que triunfe en cada caso. Por iso falamos, no peor dos escenarios, de «erros gramaticais» (i.e., erros que son en realidade gramaticais).

Considérense, a título de exemplo, as seguintes construcións, propostas polo autor, en que se dá unha dupla posibilidade de concordancia verbal, con alternancia de número (os signos [+ / ++] marcan graos crecentes de custo, peor aceptabilidade e carácter menos canónico das escollas):

- (18) a. Un grupo de persoas votou / +votaron en contra da proposta
- b. Unha chea de persoas +votou / votaron en contra da proposta
- c. Un millón de persoas ++votou / votaron en contra da proposta

A variabilidade na concordancia de número no *target* (i.e., no verbo) vén inducida polo desaxuste forma/contido no controlador, que ten nos tres

casos idéntica estrutura formal: FN co esquema [*un N de N*]. A covariación de trazos formais (en singular) é habitualmente considerada a forma de concordancia canónica e menos custosa, como ocorre en (18a); porén, en (18b) deixa de selo e máis aínda en (18c), que sería, abofé, a solución adoptada por moi poucos. Nas condicións deste último exemplo hai unha clara preferencia pola covariación entre o contido do controlador e a forma do *target* (en plural). As escollas non responden a ningún criterio formal (a estrutura frásica do controlador é idéntica nos tres casos), mais tamén non son inocuas nin inmotivadas: reflicten o diferente peso cognitivo que na mente do falante ten en cada caso a entidade controladora da concordancia, concibida ora como un colectivo definido (i.e., como un grupo), ora como unha acumulación de individuos (i.e., como persoas nunha (in)determinada cantidade).

A variabilidade nas ocorrencias de concordancias conflitivas, produto dalgunha incongruencia no controlador, poden afectar en galego a tres categorías: xénero, número e persoa. Aínda que a problemática de desaxuste que presentan e as solucións de reaxuste a que dan lugar son substancialmente comúns, o grao de «problematicidade» potencial de cada unha delas é moi diferente. Velaí os contundentes datos cuantitativos que nos fornece ao respecto *CORTEGAL*¹⁴:

— Problemas de selección inadecuada de número :	854	73,69%
— Problemas de selección inadecuada de xénero :	263	(22,69%)
— Problemas de selección inadecuada de persoa :	42	(3,62%)
TOTAL	1159	(100,00%)

Nas seguintes seccións analizaremos cada un destes tres escenarios de escolla problemática do morfema concordante, na procura das chaves diferenciais de cada unha delas.

¹⁴ Téñase en conta, porén, que neste corpus non se anotan os casos en que existe variabilidade máis ou menos problemática senón «*só o que claramente é non estándar (no seu sentido amplo)*». Só se etiquetarán como problemas aquelas formas ou usos que sexan claramente incorrectos ou non estándar, ou que sexan claramente inadecuados desde o punto de vista discursivo. Non se etiquetarán como erradas aquelas formas que son estándar aínda que non recomendadas» (*Manual de anotación en CORTEGAL*). Considérese tamén, en sentido inverso, que cando un mesmo problema de concordancia ten manifestacións múltiples en diferentes *targets*, compútase unicamente unha vez, mentres que as restantes se etiquetan como «correccións derivadas» da primeira, non como erros de seu. Por tanto as cifras dadas refírense ao número de problemas de selección inadecuada e claramente non estándar, non ao número total de veces en que eses problemas se manifestan. Para máis detalles *vid.* Álvarez de la Granja e Rodríguez Rodríguez, §2.5.5 neste volume.

6.2. Problemas de selección inadecuada de número

Os problemas de selección inadecuada de número son, con moita diferenza, os máis frecuentes, pois achegan, como se acaba de ver, practicamente tres cuartas partes dos erros de concordancia etiquetados no corpus. Constitúen tamén o erro gramatical máis frecuente e, en termos absolutos, un dos erros de lingua máis detectados no corpus, avantaxado só polos erros de puntuación e de acentuación (no plano ortográfico) e polas escollas de léxico non estándar. Tan alta presenza parece relacionarse coa grande importancia que a categoría número ten na conceptualización das entidades e dos eventos e a conseguinte existencia de numerosos procedementos gramaticais que a envolven e constrinxen, como a cuantificación ou a coordinación, por unha banda, e por outra coa forte motivación semántica da categoría (asociada de maneira moi robusta a un contido de cantidade: [SG. ‘unha entidade’ / PL. ‘varias entidades’]), que non se dá e.g. no caso do xénero, categoría moito menos motivada. Isto outórgalle certa rixidez formal a unha categoría que é, asemade, omnipresente e de uso ineludible, o cal xera numerosos desaxustes ao usala. Veláí uns exemplos tirados do noso corpus (sublíñanse os controladores):

- (19) Actualmente, a xente os fins de semana **acuden** a restaurantes, xa sea porque poden ou por simple comodidade
- (20) A maioría dos adolescentes **seguen** certas modas de marcas moi custosas que moitos pais non se poden permitir
- (21) A este tipo de xente é moi fácil recoñecela xa que cando **coñecen** o oficio ao fondo, non **dubidan** en abandonar
- (22) Desde fai un par de anos, o mundo da cociña e da gastronomía **acadaron** moita importancia
- (23) A compra de produtos de marcas brancas ou de baixa calidade si que nos **axudan** a aforrar mais cartos
- (24) Esta nova demanda de alimentos sanos, combinada coas ansias de novidades, **abiron** paso a innovación e expansión culinaria
- (25) Actualmente, o consumo e a produción de bens **é** a base de toda sociedade medianamente modernizada e desenvolta

Nos máis dos casos, a escolla inadecuada de número vén inducida por unha incongruencia entre a forma e o contido do controlador da concordancia, case sempre un nominal con morfoloxía singular que designa unha pluralidade de entidades. O formato dos desaxustes pode variar, por exemplo (a relación que segue non tenta ser exhaustiva, aínda que si significativa):

- a) A forma máis canónica —pero non a máis frecuente— desta incongruencia é a lexicamente inherente aos denominados nomes colectivos, como *xente* en (19) ou *poboación* en (7).

- b) Máis abundosa en solucións non canónicas é a construción cuantificadora [det. N_1 de N_2] de controladores como *a maioría dos adolescentes* (20) ou *unha parte dos presentes* (8), que agrupa como un colectivo N_1 individuos da clase N_2 (aínda que tamén se pode aplicar a N_2 que de seu eran xa colectivos, como acontece en (21)). Nestes casos a figura do controlador escíndese en dous, que inducen concordancias diverxentes entre si: N_1 induce unha concordancia en singular (considerada canónica por ser el formalmente o núcleo da frase), mentres N_2 desencadea unha copia de trazos semánticos e daquela un *target* en plural (os controladores secundarios aparecen subliñados con trazos descontinuos). Existe unha ampla aceptación social de ambas as alternativas, de xeito que ocorrencias como as de (20) non aparecen etiquetadas como erro en *CORTEGAL* (por suposto, tampouco o estaría a de (8), pero este exemplo, como se indicou, non está tirado deste corpus).
- c) Menos aceptadas e tamén relativamente menos frecuentes son as escollas non canónicas de concordancia na construción non-cuantificadora homomorfa da anterior [det. N_1 de N_2], que exemplifican (22-23): o complemento do nome N_1 (i.e., N_2) capta en ocasións talmente a atención do falante que asume o rol de controlador da concordancia, privilexio que formalmente non lle corresponde a el senón a N_1 , como núcleo da frase. Na orixe desta mudanza crítica sitúase un apagamento cognitivo de N_1 similar ao que se dá en (b), pero que neste caso resulta máis custoso de asumir dado que a entidade nuclear apagada non é un nome de contido estritamente cuantificador (*maioría, parte, grupo...*), senón un de contido léxico pleno (*mundo e compra*, nos exemplos dados), que se ve invisibilizado por un controlador «intruso»: a viabilidade do enunciado *A compra de produtos de marcas brancas axúdannos a aforrar* alicérase na imaxe mental ‘Os produtos de marcas brancas axúdannos a aforrar’ (vid. *infra* §6.3 para o mesmo proceso referido á selección conflitiva de xénero).
- d) A construción sintáctica que a gramática diseña especificamente para codificar un agrupamento de entidades ([A + B]) é a coordinación (i.e., [A e B]), que induce concordancia en plural, considerada canónica. Porén, existen construcións complexas distintas da coordinación que, baixo determinadas condicións pragmáticas, inducen tamén concordancia en plural, gramaticalmente irregular, como a que se aprecia en (24), formalmente [A con B], pero que representa igualmente a imaxe mental [A + B].
- e) En (25) vemos inversamente como unha construción coordinativa (o *consumo e a produción*) desencadea unha concordancia en singular

ao ser concibida como unha entidade conxunta; unha solución non canónica inversa á anterior pola cal optan un número moi significativo, aínda que minoritario, de autores en *CORTEGAL*.

Para alén das diferenzas de formato, o mecanismo que conduce a todas estas solucións non canónicas é o mesmo: o *target* copia os trazos semánticos do controlador, toda vez que os formais resultan incongruentes por non se axustaren ao padrón semántico que motiva a categoría. A toma de decisión está cognitivamente inducida polo maior peso conceptual que na mente do emisor asume o (constituínte do) controlador que restaura o equilibrio perdido: a conceptualización da *xente* como ‘persoas’, máis que como unha masa uniforme en (a); a prevalencia mental de N_2 sobre N_1 , á marxe da posición formal que ocupa na frase, en (b-c); a preeminencia que se outorga, en (d), a entidades que, aínda que situadas fóra do núcleo frásico, se conciben como agrupadas coas que si ocupan esa posición; a conceptualización integrada de varias cousas vistas como unha, en función da forte cohesión que presentan nunha determinada situación comunicativa, como en (e)...

Existen múltiples manifestacións adicionais do ascendente cognitivo na escolla. A conceptualización plural dos nomes colectivos, e.g., verifícase case sempre con nomes que designan colectivos de persoas (*a xente, a directiva, o alumnado, o profesorado*), moi raramente de animais (*‘O rabaño... andan polo monte*) e case nunca de cousas (*‘‘‘A cristalaría... son moi fráxiles*), antropocentrismo que pon de manifesto estarmos perante un mecanismo rexido pola nosa maneira de concibir a realidade e de construír o significado, non polas propiedades formais da gramática. Noutra orde de cousas, no escenario (e) antes descrito, por exemplo, é relevante constatar como as ocorrencias de conceptualización integrada de entidades coordinadas, que conduce neste caso a unha concordancia non canónica en singular, se concentran no noso corpus en coordinacións de entidades que constitúen o tema do texto (*o consumo e a produción; a gastronomía e a cociña*), que está inducido na propia tarefa, o que facilita a súa conceptualización unitaria¹⁵. Cada exemplo pide a súa reflexión crítica, i.e., a súa análise, pero todas se canalizan polos mesmos vieiros.

¹⁵ Os integrantes das coordinacións e a centralidade destas explícitanse xa no enunciado das preguntas do exame: «Redacta un texto sobre a importancia que teñen o consumo e a produción (ou o consumismo e a produtividade) no noso modo de vida actual»; «Nos últimos anos a gastronomía e a cociña acadaron moita popularidade. Redacta un texto expoñendo a túa opinión sobre este fenómeno». Repárese, porén, en que os enunciados están redactados coa concordancia gramatical canónica, en plural, o que non obsta para que unha parte significativa das persoas redacten os seus textos reconvertendo a concordancia ao singular.

6.3. Problemas de selección inadecuada de xénero

Os problemas de selección inadecuada de xénero son ben menos frecuentes que os vinculados ao número, pero relativamente abundantes, pois achegan máis dunha quinta parte dos erros de concordancia marcados no corpus. Esta correlación parece vincularse co menor peso que a categoría gramatical xénero ten na conceptualización das entidades en xeral (e en particular na das non sexuadas e non humanas, que son a maioría) e, conxuntamente, coa feble motivación semántica da categoría, que con moita facilidade se desvincula da contraposición de SEXO [MASC. ‘entidade macho’ / FEM. ‘entidade femia’]. Aínda que no caso das entidades humanas esa correlación é moito maior e adquire un importante simbolismo social, o seu manexo non adoita dar lugar a desaxustes de concordancia. Vexamos uns cantos exemplos significativos de escolla inadecuada de xénero tirados de *CORTEGAL*:

- (26) Esta ‘cacería’ do máximo rendimento da produción vai **ligado** ó aumento do consumismo. Cantos anuncios se ven en todas partes e dinche que tes que probar a nova colonia, ler o novo libro, ver a nova película?
- (27) Na actualidade a xente consume productos de marca blanca ou de marca descoñecida porque teñen menos impostos e polo tanto máis **baratas**
- (28) Existen moitos debates sobre se este sistema é o adecuado ou non, a vez que existen outras alternativas como o comunismo, gran **inimiga** do capitalismo
- (29) A humillación constante a que é **sometida** o concursante polo xurado, causa moito morbo
- (30) Moitos dos adolescentes da nosa sociedade están mercando continuamente prendas [de roupa] para mostrarse sempre á última. Neste ámbito, os prezos non soen ser baixos, xa que, non son **considerados** bens de primeira necesidade
- (31) Hai países na que a vida dos seus habitantes limítase a producir para ter cartos e consumir

Na orixe destas concordancias erradas de xénero non adoita estar un controlador léxico que amose incongruencia entre forma e contido (como ocorría cos nomes colectivos —e.g. *xente*— na concordancia de número): tal podería darse unicamente no reducido grupo dos nomes epicenos cando se aplican a entidades de sexo oposto ao xénero gramatical. Porén, mesmo nestes casos non adoitan producirse escollas de concordancia conflitivas: *Antón é moi {boa/*bo} persoa e {unha/*un} testemuña {honesta/*honesto}*¹⁶.

¹⁶ No caso dos nomes epicenos de xénero masculino (e.g. *membro*, *lince*) aplicados a entidades femias, dáse unha crecente tendencia a caracterizar de maneira motivada os seus complementos na FN co xénero feminino, mais iso non constitúe un problema de concordancia anómala, senón de cambio de xénero do controlador da concordancia: trátase do resultado gramaticalmente canónico dunha tendencia a converter os epicenos de xénero masculino en nomes de xénero común, que teñen de seu ese comportamento: «É tamén unha membro activa da Rede de Mulleres na Prevención de

O que atopamos case sempre na antesala da decisión errada na concordancia de xénero é a emerxencia dun controlador inesperado, que desbanca o candidato xenuíno (nos exemplos identifícanse cun subliñado descontinuo e continuo, respectivamente). Na imaxe mental que o falante constrúe da predicación aparece unha figura á cal finalmente non lle outorga, no enunciado, o privilexio sintáctico de ser o núcleo de frase e por tanto de gobernar a concordancia; porén, no curso do proceso de construír o enunciado esa figura adquire inopinadamente unha preeminencia conceptual que fai que o falante acabe por escollela como controlador, a pesar de non cumprir os requisitos formais para selo... porque non llos concedera previamente. Estamos diante dun trasacordo sobrevido moi común e tolerado na lingua oral, pero xulgado como moi pouco apropiado na escrita, á que se lle atribúe unha dinámica demorada e reflexiva que debería evitar tales inconsistencias.

Reparemos nas reformulacións mentais que dan lugar ás concordancias anómalas dos exemplos que se acaban de presentar:

Quen escribe (26) concibe que aquilo que *vai ligado ao aumento do consumismo* é directamente *o máximo rendimento*, non a súa *'cacería'*.

En (27) acaba por prevalecer a imaxe das *marcas brancas e descoñecidas* (máis que a dos seus *produtos*) como *máis baratas*.

En (28) quen escribe concibe que a *gran inimiga do capitalismo* non é tanto o *comunismo* como a *alternativa* comunista (i.e., o *comunismo* concibido como *alternativa*).

En (29) á idea do *concurante* sometido sobrepónselle a *da humillación constante* como sometedora (aínda que literalmente se di que é *sometida*)

En (30) a prevalencia cognitiva que acadan os *prezos non-baixos* fai que sexan estes (e non as *prendas* que teñen tales prezos) os que son *considerados bens de primeira necesidade*.

O normal é que o controlador «intruso», que capta a atención do falante ao punto de levalo a reformular a construción sintáctica mesmo incorrendo en anacoluto, estea localizado na cadea lingüística antes do *target* reformulado, pois esa precedencia habilita a reformulación do elemento subsecuente. Porén, non faltan exemplos como (31) en que aparece posposto. Quere dicirse que o controlador intruso (*a vida dos seus habitantes*) despraza o xenuíno (*países*) do seu rol e induce a modificación do *target* (*nos que > na que*) antes da súa realización fónica... pero despois, obviamente, de adquirir a súa presenza protagonista na imaxe mental do falante, que constitúe sempre o alicerce semántico sobre o cal se asenta e levanta a construción sintáctica.

Conflitos» (https://gl.wikipedia.org/wiki/Flaviana_Charles). «Morren os cachorros de 'Nenúfar', a lince morta a tiros por un cazador» (<https://www.gciencia.com/fotonautas/morren-os-cachorros-de-nenufar-a-lince-morta-a-tiros-por-un-cazador/>).

Existen contextos sintácticos que, pola súa propia natureza, propician a ocorrencia de selección non canónica de xénero, como vimos tamén coa categoría número. Entre elas non está particularmente o contexto [N_1 de N_2], a pesar de ocorrencias como (26-27): non debe confundirse coa isomorfa construción cuantificadora, que si constitúe un contorno favorecedor das concordancias non canónicas de número, como foi visto. Si se atopan entre eses contextos propicios as construcións coordinativas (que tamén o eran para o número) e as distributivas recíprocas do tipo [*un(ha) - outro(/a)*], que xeran situacións problemáticas de concordancia de xénero cando os termos empacitados na construción son nominais con xéneros distintos (en *CORTEGAL* son moi frecuentes as ocorrencias do tipo [*o consumo - a produción*], dado que é o tema dunha das redaccións que integran o corpus). Dado o funcionamento da gramática do galego, que impide esquivar o dilema, a solución pasa por escoller un único morfema, sexa o masculino ou o feminino, para recuperar todos os nominais, sen poder distribuílos segundo o seu xénero específico. As ocorrencias amósannos un panorama repartido entre a selección do masculino e do feminino dun xeito aparentemente equilibrado e aleatorio, que pode mesmo actuar nun mesmo enunciado en sentido inverso cando a operación se repite. Iso é o que ocorre en (32), onde atopamos unha construción coordinada con diversidade de xénero entre os seus membros, que son retomados referencialmente mediante dúas operacións coordinadas entre si, unha resolta coa concordancia en masculino e a outra en feminino:

- (32) En conclusión, a produción e o consumo están **interrelacionados** e unha depende da outra

Cómpre unha análise máis profunda do fenómeno para chegar a coñecer as motivacións reais das escollas, que terán plausiblemente un fundamento cognitivo.

6.4. Problemas de selección inadecuada de persoa

A concordancia de persoa é, con moito, a que menos problemas de selección inadecuada de categoría produce. En *CORTEGAL* achega, como se indicou, unicamente un 3,62% do total de ocorrencias conflitivas de concordancia anotadas. Os exemplos, ademais de escasos, teñen un perfil bastante definido:

- (33) Nos, a poboación de este mundo, necesitar **necesita** seguir alimentándose
 (34) Cando poñemos a televisión rara vez é pra ver un documental, por exemplo.
Buscan entretemento
 (35) Somos unha sociedade insaciable que **queremos** aquilo que vemos

- (36) Producir a través do consumo de outros un beneficio para ti, no que **tú** gastas para comer, **vestirse**, mercar un xoguete... (consumir)
- (37) A publicidade está en todas partes e non **somos** ninguén se non **tes** un «iphone» na man e **vistes** unhas «adidas»

É sabido que a categoría persoa serve para identificar, na morfoloxía verbal e pronominal, as instancias responsables da interacción comunicativa: ‘quen fala/escribe’ coa «primeira» persoa (P1); ‘a quen se fala/escribe’ coa «segunda» (P2); para alén da «terceira», a non-persoa coa que se designan as outras entidades que aparecen referidas no discurso (‘de quen se fala/escribe ≠ P1/P2’). O manexo do sistema raramente dá problemas, sobre todo cando P1 e P2 fan referencia explícita aos intervenientes no acto comunicativo como participantes da predicación. Os escasos problemas acostuman ocorrer, como se pode verificar nos exemplos, cando aparecen como produto de estratexias de (inter)subxectivización (no senso de Traugott, 2010) e non como participantes de seu na predicación. A persoa que fala ou escribe pode optar por unha estratexia de implicación no contido da mensaxe, presentándose como partícipe daquilo que di (usando a P1 pl., e.g. *a xente nova facemos X*) ou implicando o oínte/lector na procura da súa solidariedade (usando a P2: *se es novo/a fas X*), fronte a unha estratexia obxectivista, de distanciamento con relación ao contido da mensaxe (usando a P3: *a xente nova fai X*) (vid. ao respecto o capítulo de Silva-Domínguez neste mesmo volume). É unha decisión estratéxica e de perspectiva. Neste caso non son raros os problemas de concordancia por falta de congruencia na decisión: i.e., como produto de mesturar ou de cambiar de estratexia ao longo do texto e mesmo dentro dos límites do enunciado. En (33) cabe unha estratexia subxectivizada (*Nós, a poboación deste mundo, necesitamos...*) ou de distanciamento obxectivo (*A poboación deste mundo necesita...*); o que non cabe é mesturalas de xeito incongruente. Razoamentos similares serven para explicar as escollas inadecuadas de número nos restantes exemplos.

Particular atención reclama o crecente uso da P2 (*ti*) cun aparente valor xenérico ou impersoal e produto dunha estratexia de solidariedade, polo xeral tamén aparente, coa persoa interlocutora, como o que (con incongruencia incluída) se exemplifica en (36) e (37) (vid. tamén Silva Domínguez nesta obra). Trátase dun uso que se sitúa nunha interesante intersección entre as tendencias evolutivas cara á subxectivización e á informalización (tamén chamada demotización) da linguaxe, que se verifica con este mesmo ítem en diferentes linguas (Skärland, 2017). En canto á segunda liña, resulta aínda dominante en galego (fronte a outras linguas como o inglés, e.g.) a

consideración destes usos como inadecuados para os rexistros formais, entre os que cabe incluír a escrita académica. A alta frecuencia con que atopamos no corpus estes usos serían atribuíbles daquela a unha escolla inadecuada de rexistro; mais podería ser tamén, finalmente, unha mostra da tendencia á informalización do discurso público no mundo occidental post-industrial, coa incorporación de trazos característicos da lingua coloquial, íntima ou privada ao espazo público (Fairclough, 1996; *vid.* Nordquist, 2020 para unha visión xeral introdutoria da cuestión). Mais este é xa un tema que transcende o problema da concordancia, co cal só tanxencialmente se vincula, en casos en que a estratexia discursiva non se aplica de xeito coherente.

7. REMATE

Do exposto nas páxinas precedentes conclúese, entre outras cousas, que resulta inadecuado concibirmos a concordancia (nomeadamente a verbal) como un mecanismo de copiado de trazos formais dunha entidade léxica noutra. De funcionar así o mecanismo, carecerían de explicación non só todas as concordancias problemáticas e as consideradas abertamente erradas, senón moitas das canónicas. Por exemplo: de onde copiaría o verbo a marca morfolóxica de número (plural) nunha ocorrencia como *O neno e a nena marcharon?* Ou a de (primeira) persoa en *Os galegos temos lingua propia?* Obviamente, de ningunha das FN controladoras do mecanismo. Como foi sinalado, na concordancia o verbo pode copiar propiedades formais do controlador... ou propiedades semánticas; e podemos dicir que adoita copiar as dúas asemade, porque polo xeral unhas e outras son congruentes. Ora ben, en caso de desaxuste entre elas (que é o que se dá na maior parte das concordancias non canónicas), a tendencia *natural* das linguas é a de copiar os trazos semánticos, pois tende a considerarse menos custosa a incongruencia de formas (que son só un instrumento para acadar un fin comunicativo) que a de contidos (cuxa transmisión constitúe o fin mesmo da comunicación).

Esta concepción da concordancia baseada no contido correspóndese, a nivel global, coa da lingua como un sistema semántico: un sistema para construír significados coa axuda doutros sistemas formais que codifican e permiten transmitir eses significados a outras persoas (Halliday, 1985). A lingua é un «potencial de significado» que se substancia en diferentes sistemas de opcións, entre as cales os falantes realizan escollas, procurando as construcións máis axeitadas en cada caso. Así as cousas, a análise sintáctica debe consistir en explicitar eses procesos de decisión: reflexionar acerca dos motivos por que os falantes escollemos unhas opcións e rexeitamos outras,

atendendo sempre á adecuación de cada unha delas ao contido da mensaxe que queremos transmitir (adecuación semántica), ao conxunto do texto en que se integra e en que consiste a mensaxe (adecuación textual ou discursiva) e ao contexto comunicativo en que queremos transmitir esa mensaxe (adecuación pragmática). Todas as persoas facemos de maneira implícita esa análise sintáctica cada vez que falamos; e facémola case sempre ben.

A concordancia, obviamente, tamén forma parte dese xogo de encrucilladas. O sistema lingüístico obríganos a escoller entre ese mecanismo e outros no mercado das relacións gramaticais; e unha vez que optamos por el, o seu manexo consiste nunha constante selección de opcións. Estas habitualmente están moi automatizadas e pautadas, polo que resultan practicamente invisibles para o usuario. No caso de existir desaxuste entre a forma e o contido das opcións, as escollas diversifícanse e as solucións, sexan cales sexan, tórnanse visibles a forza de seren problemáticas e mesmo controvertidas. Esa desautomatización, visibilización e mesmo controversia son unha beizón para estimular a análise sintáctica.

Os exemplos problemáticos ou «errados» que se nos presentan decote (nas conversas, nas lecturas, na aula... en *CORTEGAL* mesmo!) son pezas que non podemos desaproveitar como recurso didáctico para sometelas a unha análise reflexiva, convenientemente encamiñada, na aula. Esa análise, como toda análise sintáctica, ten que estar baseada na intuición que todos e todas temos como falantes e asemade na formación argumentativa en que ten que consistir a aula de gramática, xa que vimos sen ela «de fábrica». Esa reflexión intelectual acerca de como nos comunicamos (incluíndo os usos problemáticos, que son sempre «nosos» porque ou ben os producimos ou ben os procesamos, case sempre con éxito) levaranos a coñecer aos poucos como funciona ese marabilloso mecanismo que é a linguaxe, sen o cal non seríamos quen somos nin como somos.

Os camiños polos que esa indagación nos pode levar son infinitos e difíciles de predicir, mais por iso mesmo sempre intrigantes. Frecuentemente sorprenderannos conducíndonos a descubrimentos imprevistos. Un só botón de mostra: imaxinemos que batemos coa peculiar concordancia dos seguintes exemplos (o primeiro é de *CORTEGAL*, os outros achéganse como proba do espallamento desta construción):

- (38) Tense habituado a unha conducta machista que é, que neste ámbito [o da moda] son mellores as mulleres, ou é propio destas **cosa que é falso**
- (39) Hai algunha conselleira non presente agora, e tamén algún conselleiro, que presume de que se lles van dar unhas pensións ás amas de casa, **cousa que é falso**, literalmente falso (Diario de sesións do Parlamento de Galicia, 23.12.1999)

- (40) En primeiro lugar dicir que a alegación da telefonía móbil basease en que se crea unha taxa polo 1,5%, **cousa que é falso** porque é pola utilización do dominio público local (Sesión plenaria do Concello de Pontevedra, 21.12.2007)
- (41) Nós dicímoslle a vostede, **cousa que é certo**, que fan estradas onde non hai coches (Diario de sesións do Parlamento de Galicia, 17.11.2000)

Como xustificar esta escolla de concordancia en masculino, aparentemente agramatical? Non existe desaxuste entre forma e contido no controlador, de xénero feminino inmotivado, nin atopamos outro constituínte masculino que destaque na conceptualización das mensaxes e emerxa como controlador alternativo. Pode parecer así (38) o produto illado dalgún tipo de desleixo, mais o certo é que, rastrexando no caso, verificamos unha construción recorrente e de aparición relativamente sistemática mesmo en rexistros formais como a linguaxe parlamentaria (39-41). O certo é que a análise dunha concordancia anómala aparentemente illada condúcenos a descubrir, de maneira inopinada, un proceso de gramaticalización en avanzado curso, no cal se está a fixar un serio candidato a novo nexos de oracións relativas (*cousa que*, concorrendo con *o que*).

Unha análise sintáctica como a que aquí se propón pode xurdir na aula de maneira espontánea, a raíz da aparición deste ou aquel uso controvertido da concordancia, que merece ser comentado. Aínda que sen a vantaxe da naturalidade, unha análise máis estruturada pode artellarse, como unha tarefa dirixida, a partir dunha escolma de exemplos convenientemente recollidos (o que ten a vantaxe alternativa de podermos traballar unha casuística diversa, representativa do conxunto de problemas envoltos); ou tamén a partir da análise dun texto real (o que ten a vantaxe de situar a análise no marco textual, que é o máis acaído para o caso). O texto que reproducimos como (42) podería ser un bo candidato polas perspectivas de reflexión crítica que abre. Trátase do texto CORT_0925 do noso corpus:

- (42) Hoxe en día, nas nosas casas, dispoñemos de televisión máis internet. Podemos consumir isto varias veces ao día con: películas, partidos de fútbol, series... É a forma que teñen a infancia e a mocidade de entreterse. Ao estar vindo e seguindo de cerca, por exemplo, os partidos de fútbol fai que os idolatren e queiran ser coma eles.

Co paso dos anos a xuventude cada vez seguen a mais xente e queren ser coma eles. Isto sucede, principalmente, polas televisións. Cando poñemos a televisión rara vez é pra ver un documental, por exemplo. Buscan entretemento. A televisión xa véndenos a parte boa da situación como, por exemplo, un fichaxe de fútbol por millóns de euros ou un desfile dun modelo no que gaña moito diñeiro. A xuventude ao ver isto, queren chegar a ser coma eles e deixar de la moitas cousas que sí son positivas para a súa vida.

En conclusión, a xuventude déixase influenciar moito polo que ve en televisións ou revistas sen coñecer a súa parte perxudicial e non é capaz de

apreciar realmente todo o esforzo que poden ter un escritor, por exemplo. Penso que nas televisións deberían de facer máis incapé nestas cousas para concenciar máis aos nenos.

Non vou desenvolver un pautado miúdo da actividade, que debe ser flexible para adaptarse á dinámica xerada na aula en cada caso e que admite numerosas variacións (*vid.* e.g. Romero et al., 2018 para unha proposta similar, no ronsel tamén dun estudo sobre competencia gramatical na redacción de textos escritos por alumnado arxentino de primeiro ano de universidade).

A análise podería iniciarse coa proposta de identificar as concordancias anómalas ou problemáticas presentes no texto (tarefa que facilmente se verá interferida polo sinalamento de anomalías que correspondan a fenómenos distintos da concordancia, que cómpre deslindar) e o debate acerca da súa aceptabilidade, no que é previsible que se evidencie certa falta de consenso.

A continuación, para cada ocorrencia problemática identificada deberanse indagar e expoñer, de xeito argumentado, os posibles motivos da escolla feita pola persoa autora do texto. Identificaranse nel o controlador ou controladores que interveñen na decisión, analizándoos na procura dalgún desaxuste entre forma e contido, así como a eventual existencia de controladores «intrusos» aos que formalmente non correspondería ese rol mais que captan a atención do falante ao punto de levalo a reformular a construción sintáctica dun xeito formalmente inadecuado.

En terceiro lugar, para cada ocorrencia problemática débense analizar as opcións de concordancia alternativas existentes, indagando na adecuación semántica e pragmática que amosan para o caso concreto ou para outros en que puidesen ser utilizados.

Finalmente, farase unha avaliación global do uso da concordancia no texto, axuizando a maior ou menor repercusión das escollas problemáticas ou erradas na coherencia textual, incluíndo a coherencia da estratexia de posicionamento do emisor na construción do seu propio discurso. Farase tamén unha reavaliación da aceptabilidade dos usos problemáticos (xustificando posibles mudanzas de criterio verbo da inicialmente feita) e, como conclusión, proporase unha reescritura do texto coas formas (de concordancia, pero tamén doutro tipo) que se consideren óptimas para obtermos un produto adecuado á situación comunicativa¹⁷.

¹⁷ Aínda que é obvio que ese texto final pode presentar formas diversas, cabe lembrar que en *CORTEGAL* se ofrece, para todos os textos recollidos no corpus, unha proposta de reescritura estándar, disposta de maneira estratificada e acumulativa en estándar ortográfico, morfolóxico, léxico, gramatical (i.e., sintáctico), semántico e discursivo, cuxo manexo pode ser de utilidade.

En definitiva, trátase de desenvolver unha análise baseada na reflexión e na avaliación argumentada de opcións. Unha análise que evoque, explicitándoa, aquela que implicitamente todas as persoas facemos cada vez que falamos ou que escribimos. Unha «análise con análise», que é tanto como dicir unha análise con sentido; en todos os sentidos da palabra. Porque tendemos a pensar que a análise sintáctica é un exercicio arbitrario que non ten sentido... e tendemos a concordar co que pensamos. Pero non sempre pensamos adecuadamente as cousas e, ás veces, máis que xuízos construímos prexuízos, no que é unha forma particularmente daniña e cativadora de conforto intelectual. Por fortuna, sempre podemos darlles unha volta aos nosos pensamentos e aos nosos (pre)xuízos. Neste senso, a análise dos erros de concordancia pode sernos, con certeza, de moita axuda.

REFERENCIAS BIBLIOGRÁFICAS

- ASTOLFI, J. P. (1997). *L'erreur, un outil pour enseigner*. ESF (Trad. esp.: *El «error», un medio para enseñar*. Díada. 1999).
- BARLOW, M. e FERGUSON, Ch. A. (Eds.) (1988). *Agreement in Natural Language: Approaches, Theories, Descriptions*. Center for the Study of Language and Information, Stanford University.
- BOSSONG, G. (1980). *Aktanten Funktionen in romanischen Verbalsystem*. *Zeitschrift für romanische Philologie*, 96(1-2), 1-22. <https://doi.org/10.1515/zrph.1980.96.1-2.1>
- BRESNAN, J. (1995). Category mismatches. En A. Akinlaby (Ed.), *Theoretical Approaches to African Linguistics* (pp. 19-46). Africa World Press.
- BROWN, Ch. e SEMPERE, J. (1987). Los clíticos del español vistos como inflexiones. *Revista Argentina de Lingüística*, 3(1), 69-83.
- CIDRÁS, F. (1992). Sobre a función da concordancia. Consideracións a propósito das concordancias verbais anómalas. *Verba*, 19, 41-53.
- CIDRÁS, F. (2009). Os camiños da gramática: unha aproximación á variación e o cambio lingüístico desde a sintaxe do galego. *Estudos de Lingüística Galega*, 1, 27-46.
- CIDRÁS, F. (2012). On the concept of syntactic function in a functional grammar. *Functions of Language*, 19(2), 147-173. <https://doi.org/10.1075/fol.19.2.01cid>
- CORBETT, G. G. (2006). *Agreement*. Cambridge University Press.
- CROFT, W. (1988). Agreement vs. Case Marking and Direct Objects. En M. Barlow e Ch. A. Ferguson (Eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions* (pp. 159-179). Center for the Study of Language and Information, Stanford University.
- DE LA TORRE, S. (2004). *Aprender de los errores. El tratamiento didáctico de los errores como estrategia de innovación*. Magisterio del Río de la Plata.

- DU BOIS, J. W. (1985). Competing motivations. En J. Haiman (Ed.), *Iconicity in Syntax* (pp. 343-365). John Benjamins. <https://doi.org/10.1075/tsl.6.17dub>
- FAIRCLOUGH, N. (1996). Border crossings: Discourse and social change in contemporary societies. En H. Coleman e L. Cameron (Eds.), *Change and Language* (pp. 3-17). Multilingual Matters.
- FRANCIS, E. J. e MICHAELIS, L. A. (Eds.) (2003). *Mismatch: Form-Function Incongruity and the Architecture of Grammar*. Center for the Study of Language and Information, Stanford University.
- FRANCO, J. (2000). Agreement as a continuum: The case of Spanish pronominal clitics. En F. Beukema e D. Dikken (Eds.), *Clitic Phenomena in European Languages* (pp. 147-190). John Benjamins. <https://doi.org/10.1075/la.30.07fra>
- FREIXEIRO MATO, X. R. (2009). *Lingua de calidade*. Xerais.
- Fromkin, V. A. (Ed.) (1984). *Speech Errors as Linguistic Evidence*. Mouton. <https://doi.org/10.1515/9783110888423>
- FROMKIN, V., RODMAN, R. e HYAMS, N. (2011). *An Introduction to Language* (9.^a ed.). Cengage Learning Wadsworth.
- GARCÍA-MIGUEL, J. M. (1991). La duplicación de complemento directo e indirecto como concordancia. *Verba*, 18, 375-410.
- GILQUIN, G. e GRANGER, S. (2022). Using data-driven learning in language teaching. En A. O’Keeffe e M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (2.^a ed., pp. 430-442). Routledge. <https://doi.org/10.4324/9780367076399>
- HALLIDAY, M. A. K. (1985). *An Introduction to Functional Grammar*. Edward Arnold.
- HERNANZ, M. L. e BRUCART, J. M. (1987). *La sintaxis. 1. Principios teóricos. La oración simple*. Editorial Crítica.
- JESPERSEN, O. (1922). *Language: Its Nature, Development and Origin*. George Allen & Unwin Ltd. <https://doi.org/10.4324/9780203715895>
- MCENERY, T. e RICHARD, X. (2010) What corpora can offer in language teaching and learning. En E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 364-380, vol. 2). Routledge.
- MORTON, J. (1964). A model for continuous language behaviour. *Language and Speech*, 7, 40-70. <https://doi.org/10.1177/002383096400700105>
- NORDQUIST, R. (2020, 28 de agosto). Informalization in Language. *ThoughtCo*. <https://www.thoughtco.com/informalization-in-language-1691066>
- PERES, J. ANDRADE e MÓIA, T. (1995). *Áreas Críticas da Língua Portuguesa*. Caminho.
- RAPOSO, E. B. P., BACELAR DO NASCIMENTO, M. F., COELHO DA MOTA, M. A., SEGURA, L. e MENDES, A. (Orgs.) (2021). *Gramática do Português* (vol. III, 2.^a ed.). Fundação Calouste Gulbenkian.

- RÖMER, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. <https://doi.org/10.1017/S0267190511000055>
- ROMERO, M. C., MORÓN USANDIVARAS, M. e ANDRÍN, L. G. (2018). Un estudio sobre la concordancia en número en estudiantes de primer año de universidad. *Signo y seña*, 33, 143-156. <https://doi.org/10.34096/sys.n33.5262>
- SKÄRLUND, S. (2017). Does *you* mean *I*? Generic *du* ('you') as a case of informalization and subjectification in Swedish. *Norsk Lingvistisk Tidsskrift*, 35(1), 129-146.
- STEELE, S. (1978). Word order variation: a typological study. En J. H. Greenberg, Ch. A. Ferguson e E. A. Moravcsik (Eds.), *Universals in Human Language. IV. Syntax* (pp. 585-623). Stanford University Press.
- STURTEVANT, E. H. (1917). *Linguistic Change*. University of Chicago Press.
- STURTEVANT, E. H. (1947). *An Introduction to Linguistic Science*. Yale University Press.
- TIBERIUS, C., CORBETT, G. e BARRON, J. (2002). *Agreement: A bibliography*. [Project Agreement: An investigation into the distribution of information. Surrey Morphology Group. University of Surrey]. <https://www.smg.surrey.ac.uk/projects/agreement/bibliography/>
- TRAUGOTT, E. C. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. En K. Davidse, L. Vandelanotte e H. Cuyckens (Eds.), *Subjectification, Intersubjectification and Grammaticalization* (pp. 29-74). De Gruyter Mouton. <https://doi.org/10.1515/9783110226102.1.29>
- VÁZQUEZ ROZAS, V. e BLANCO, M. (2022). Corpus y enseñanza del español. En G. Parodi, P. Cantos-Gómez e C. Howe (Eds.), *Lingüística de corpus en español. The Routledge Handbook of Spanish Corpus Linguistics* (pp. 342-356). Routledge. <https://doi.org/10.4324/9780429329296-26>
- YUASA, ETSUYO (2005). *Modularity in Language: Constructional and Categorical Mismatch in Syntax and Semantics*. Walter de Gruyter. <https://doi.org/10.1515/9783110197525>

Procedementos de xeneralización e marcado da implicación no corpus *CORTEGAL*: entre a linguaxe académica e o rexistro coloquial da lingua

CARME SILVA-DOMÍNGUEZ
Universidade de Santiago de Compostela

1. INTRODUCCIÓN¹

O presente traballo ten como obxectivo analizar como se expresa a relación entre lugar común e punto de vista do/a autor/a nos textos que constitúen o corpus *CORTEGAL*, presentado en Álvarez de la Granja e Rodríguez Rodríguez neste volume. Estudárase o inventario de recursos que permiten tanto a expresión de ideoloxías, opinións etc., partilladas por un número indeterminado de persoas, como a asunción e expresión do parecer máis persoal, desprovisto de todo apoio colectivo. Entre ambos os extremos sitúase a ampla gama de posibilidades que permiten soste un punto de vista persoal dotándoo da aparencia de opinión maioritaria. Dada a estreita relación existente entre a diversidade de opcións estruturais e a oposición entre variedades coloquial e formal da lingua, recádaranse datos acerca da adscrición a diferentes rexistros de cada un dos procedementos detectados no corpus, coa finalidade de indagar sobre que papel desenvolve nas aulas de lingua galega o concepto *adecuación*, de tanta importancia para o proceso de construción textual.

¹ A autora do traballo quere agradecer ás profesoras Noemí Basanta Llanes e Olga Amigo Devesa a súa xenerosa achega de información sobre distintas cuestións relativas á composición de textos académicos nas aulas de Bacharelato. Os erros, como é obvio, son só responsabilidade da persoa asinante.

A procura bibliográfica de referencias para a análise do tipo de estruturas escollida permitiu localizar diversos traballos que estudan a tamén chamada *desaxentivación*, *despersonalización* e *impersonalización*, conxunto de procesos vinculados á elaboración de textos científico-técnicos e consistentes no borrado da persoa emisora para a consecución dunha perspectiva máis neutral, menos subxectiva na emisión da mensaxe, que gaña así en credibilidade. Máis recentemente son obxecto de análise os recursos obxectivadores de que a lingua dispón e que se aplican nos medios de comunicación².

De entre os procedementos estudados na bibliografía baixo o paraugas da desaxentivación, no presente traballo analízase a variación nas persoas gramaticais e mais o emprego da voz pasiva. As estruturas pasivas con *se*, reiteradamente mencionadas tamén para a consecución da obxectividade discursiva, aparecen ben representadas no corpus CORTEGAL, e en menor medida ocorren nel exemplos de voz pasiva clásica:

- (1) Por outra banda tamén *se debe criticar o consumismo*
- (2) Se ben é certo (...) que a mocidade aspira a profesións como futbolista ou modelo, *un feito como este non debería ser tido en conta* para falar do tema que se propón no documento

As estruturas (1)-(2) serán analizadas como casos de *xeneralización*, pois nelas exprésanse aquelas opinións, tendencias, puntos de vista etc., que se consideran máis comúns nun momento dado e nunha determinada sociedade para a maioría, ou un sector, da poboación. Outras formas de expresar o mesmo valor, tamén obxecto de estudo para nós, aparecen nas estruturas seguintes:

- (3) *A xente* vai seguir comprando sen escrúpulos, gastando grandes cantidades de diñeiro
- (4) O apuro é cego e a fame sen mirar, *cada un* mira polo seu e non polo ben común
- (5) Os medios comunicativos provocaron que *as persoas* se interesasen por este ámbito
- (6) Hai que ter en conta que *gran parte da poboación mundial* traballa no sector terciario
- (7) Hoxe en día, *a sociedade* non sería nada se non existise o consumismo e a produtividade
- (8) O mundo móvese grazas a iso xa que *todo o mundo* quere mellorar o seu nivel de vida

² Unha noción próxima ás sinaladas é a de (inter)subxectividade, desde a cal se estudan as expresións de incerteza en Míguez neste mesmo volume.

Nalgúns casos, a intención xeneralizadora que se percibe no uso dos recursos indicados combínase coa necesidade de clarificar en que medida a persoa emisora participa do lugar común, ideoloxía, punto de vista etc., presentado como maioritario na sociedade. Exemplos como (9) e (10) explícanse polo afán de converter comportamentos, opinións colectivas, hábitos en propios tamén de quen constrúe o texto e serán analizados baixo a denominación de *marcado da implicación*:

- (9) Semella que cando *imos* ó super, comprar as cousas o mais baratas posibles, e sinónimo de ser un bó consumidor
- (10) A publicidade está en todas partes e non *somos* ninguén se non *tes* un «iphone» na man e vistes unhas «adidas»

A carón das posibilidades construtivas ligadas á xeneralización e ao marcado da implicación, o/a emisor/a dispón dalgúns recursos que permiten distanciar o seu punto de vista persoal daquel propio de determinados sectores da sociedade. É o que ocorre en (11):

- (11) *Esas persoas* que mercan estes produtos, obxectos, máquinas ou complementos carísimos e innecesarios para sentirse ben con un mesmo, non saben que están fomentando o consumismo, un exercicio moi perigoso para a sociedade e para o mundo en xeral

Indagarase así mesmo sobre unha característica común á xeneralización e ao marcado da implicación, relevante para a percepción da diferenza entre o rexistro académico e o informal: se ben os procesos que se estudarán teñen lugar en ambos, non todos os recursos presentan a mesma validez no discurso coidado. Existen algunhas opcións estruturais compartidas polo coloquio e o estilo académico, mais outras son específicas unicamente dun deles. Este aspecto será obxecto de especial interese no noso traballo, tanto no relativo ás instancias (gramaticais, académicas) que distribúen os recursos entre rexistros, como na transmisión escolar destas decisións aos/ás usuarios/as finais da lingua.

O traballo estrutúrase nas seguintes seccións: primeiramente farase unha reflexión sobre as características dos textos do corpus *CORTEGAL* examinados desde a súa condición de producións de natureza académica, e por tanto como exemplos que permiten comprobar a asimilación, por parte do/a alumno/a, de estratexias de adecuación dirixidas a seleccionar recursos especificamente ligados á linguaxe formal. Unha segunda sección presenta un achegamento bibliográfico aos procedementos de consecución da obxectividade textual na lingua galega, que serán inventariados e sobre os cales se

achegará a información dispoñible acerca da súa adecuación, ou non, para a construción de textos formais. Un cuarto apartado describirá a presenza en *CORTEGAL* de cada unha das mencionadas posibilidades construtivas e na parte final combinaranse as conclusións con algunhas pautas relativas á implementación de posibles estratexias de ensino da adecuación que permitan unha maior claridade na delimitación entre rexistros.

2. O CORPUS *CORTEGAL* E O CONCEPTO DE *ADECUACIÓN TEXTUAL*

O corpus *CORTEGAL*, de acceso libre e aloxado na páxina web do Instituto da Lingua Galega da USC, está conformado por

1000 redaccións elaboradas como resposta á pregunta número 3 [da proba de Lingua Galega no exame de acceso á universidade], en que se solicita que a/o estudante redacte un texto, de ao redor de 200-250 palabras, en que exprese a súa opinión sobre un tema de actualidade (Álvarez de la Granja, 2018, p. 55).

O nome do corpus, *Corpus de textos galegos escritos por estudantes no ámbito académico*, sinala xa un aspecto moi relevante da súa caracterización, dado que nel se subliña como central o contexto (académico) en que os materiais que o compoñen foron producidos. A intención das/os docentes que deseñaron a pregunta do exame a partir da que se constrúe o corpus foi moi probablemente a de conseguir un produto textual de extensión determinada (pois hai unha limitación en número de palabras) e redactado no rexistro formal da lingua. Un texto persuasivo sobre unha cuestión de actualidade.

2.1. Mínima aproximación ao concepto de *adecuación textual* no ámbito hispánico e aplicación da definición aos textos de *CORTEGAL*

Os tratamentos teóricos do concepto (Fernández Smith e Escoriza Morera, 2004; Gómez Capuz, 2006) subliñan a necesidade de dotar de maior relevo no traballo escolar a unha propiedade textual que se define como segue:

La adecuación supone el conocimiento y dominio de la diversidad lingüística, debida a diversos factores (geografía, historia, grupo social, situación de comunicación, canal), se basa en la noción de elección y consiste en saber escoger, de todas estas opciones, en (sic) la solución más apropiada para cada situación de comunicación (Gómez Capuz, 2006³).

³ A versión do artigo que se localizou carece de paxinación, por causa do seu formato exclusivamente dixital.

De acordo con esta achega, a adecuación pode definirse como a habilidade de escoller estruturas lingüísticas axeitadas en función de factores externos ao propio discurso. Nalgunhas aproximacións, a adecuación supón unha subpropiedade dentro da propiedade textual, máis xenérica, da coherencia:

La coherencia tiene que ver con el interior del texto pero tiene que ver, también, con la pertinencia entre texto y contexto: incoherente es el texto que no responde adecuadamente al sentido que tiene, a lo que se espera de él en el contexto en el que se produce (Côrte Vitória, 2018, p. 81).

Nos traballos examinados non se adoita defender que a oposición formal/informal estea baseada en principios absolutos e inamovibles:

La adjudicación de formalidad o informalidad a un rasgo verbal depende de los usos lingüísticos que haga de él su comunidad de hablantes. (...) De este modo, no existe nada neutro. Cualquiera palabra, construcción o giro carga con sus propias marcas, con sus connotaciones, con la historia de los usos lingüísticos que de ellos ha hecho la comunidad hablante. Por otra parte, la carga formal/informal de la lengua no es un valor estable ni absoluto para todos los usuarios (Cassany, 1995, p. 188).

Segundo se indicou anteriormente, nos materiais que conforman o corpus *CORTEGAL* a situación de comunicación en que se constrúen os textos é moi concreta: trátase do «ámbito académico», a práctica e avaliación da construción textual nun contexto escolar. Porén, indagando algo máis acerca das circunstancias de composición das redaccións, pódese detectar un certo paradoxo, xa sinalado en traballos previos sobre o corpus (Álvarez de la Granja, 2020a e b). O equipo de correctoras/es que se ocupa de achegar unha cualificación numérica á proba de lingua galega para o acceso á universidade constitúese como interlocutor primario dos materiais e, dado que a diagnóstico achegada por este terá relevancia no futuro do/a aprendiz/a, enténdese que se elaborará un texto dotado das «mellores galas» lingüísticas de que dispoña; trátase, ademais, dunha composición escrita, polo que admite reformulacións e revisións en que se implementen modificacións, sen que os borradores previos teñan presenza ningunha no proceso de cualificación. No entanto, outras circunstancias relativas ao modo de produción do texto (a ausencia dun coñecemento previo do tema e mais o tempo limitado do que a persoa dispón para a elaboración dunha redacción que se insire nun exame de varias preguntas) poderían ter como consecuencia unha redución da capacidade, por parte do/a aprendiz/a, para a realización do proceso de cribado de materiais que implica a adecuación textual. A propia situación mental da persoa produtora, derivada do momento de tensión e da transcendencia

da proba, sitúase tamén na liña de propiciar un menor control sobre o material que se constrúe.

2.2. O tratamento escolar da adecuación textual: textos legais, libros de texto e instrucións para a elaboración e a corrección do exame de acceso á universidade

O exercicio do que se toman as redaccións que conforman o corpus *CORTE-GAL* vén formando parte do exame de lingua galega para o acceso á universidade desde finais da primeira década do século XXI e tiña, no momento de toma dos materiais, un destacado papel na cualificación da proba (o 30% da nota). A actividade proposta sitúase no horizonte da produción de materiais de natureza argumentativa: a persoa que redacta debe achegar a súa visión motivada a respecto dun proceso, tendencia, comportamento ou ideoloxía que se consideran compartidos por un sector da cidadanía. Alén de medir a capacidade de defender a posición propia aducindo razóns de natureza diversa, frecuentemente sometidas nos manuais a unha tipoloxización, os criterios de corrección do exame para o curso 2016-2017, en que se obtiveron as redaccións que conforman o corpus, permiten recoñecer a vinculación destas co concepto de *adecuación textual*, explicitamente aludido:

Valórase a capacidade de construción de textos adecuados á situación (formal) en que estes se producen. Requírese un apropiado manexo dos rexistros da lingua, tanto nas escollas lexicais como no emprego das construcións gramaticais. Deben evitarse tanto os trazos propios da oralidade e dos textos espontáneos e non planificados como os que amosan un requintamento afectado, así como as palabras baúl. Bo manexo dos recursos cohesivos, tanto gramaticais como léxicos, é dicir, evitar as repeticións mediante o emprego dos recursos de cohesión. (Criterios de corrección do exame de Lingua Galega, curso 2016-2017, recuperables en <https://ciug.gal/gal/abau/exames> [última consulta: 31/01/23]).

É certo que no parágrafo reproducido se indica de maneira explícita que a diferenciación entre rexistros non aparece ligada unicamente ao compoñente léxico; porén, menciónanse a modo de exemplo erros ligados fundamentalmente ao compoñente léxico da lingua (palabras baúl, repeticións)⁴, o que podería desequilibrar a balanza do traballo escolar sobre a adecuación cara a este nivel de análise e talvez diminuír, ou colocar nun plano menor, a práctica de procesos de selección entre rexistros vinculados a outros niveis de análise. Non dispoñemos de materiais para examinar e, no seu caso, aseverar

⁴ Aínda que no segundo caso se propón un procedemento que implica o compoñente sintáctico para solucionar a dificultade orixinal (a repetición de palabras), esta é, de base, de natureza léxica.

en firme esta hipótese, mais si é posible achegar algún indicio ao respecto. Así, no libro de 1º de BAC da Editorial Anaya (Nieto Pazó e Troitiño Fráiz, 2015) exemplifícase a adecuación mediante dous fenómenos lingüísticos moi concretos, as formas de tratamento e as escollas léxicas:

(...) Non resultaría oportuno utilizar *expresións xocosas* nun funeral; non resultaría apropiado dirixirse a unha persoa maior, coa que non temos moito trato, utilizando apelativos cariñosos; ou non resultaría aceptable utilizar *termos coloquiais* nunha disertación científica [as cursivas son nosas].

A vinculación entre adecuación e procesos de selección léxica en que intervén o fenómeno da sinonimia parcial, sendo que unha das formas pertence con claridade ao rexistro formal da lingua e outra ou outras ao coloquial, vese referendada neste manual mediante o exercicio que se achega para a asimilación do concepto; nel, a/o aprendiz/a debe asociar cada unha das palabras dunha serie a unha das tres variedades lingüísticas previamente definidas (coloquial-vulgar, estándar e culta). Deste xeito sublíñase a diferenza, en termos de adecuación, entre formas próximas semanticamente do tipo de *falar, latricar, baduar* ou *parolar*, o que desde o noso punto de vista reforza, na mente da persoa que aprende, o vínculo entre o concepto de adecuación e o compoñente léxico da lingua.

Dous libros de texto máis de 1º de BAC da época LOMCE foron obxecto de análise, aínda que neles a presenza do concepto *adecuación* é notablemente máis reducida do que no anterior. Iglesias Álvarez et al. (2014) asinan o manual da Editorial Xerais. Dotado dunha estrutura peculiar, pois parte xa de inicio dunha tipoloxía textual que se estende en unidades sucesivas ata abranguer a totalidade dos temas de lingua do manual, non foi posible localizar nel actividade ningunha vinculada á práctica da adecuación; porén, varias propostas prácticas (por exemplo, o exercicio 5 da p. 22 e o exercicio 8 da p. 140) traballan a cohesión. O segundo dos manuais, publicado pola editorial Baía (González Avión e Ínsua López, 2016) tampouco contén proposta de definición ningunha para o termo *adecuación*; neste manual, unicamente un exercicio de carácter práctico aparece vinculado ao proceso (p. 155) e consiste nunha redacción para a que se achegan pautas temáticas ao tempo que se indica que se debe revisar «a súa coherencia, cohesión, adecuación e corrección ortográfica e gramatical».

A consulta dos materiais elaborados para a corrección e preparación da pregunta 3 e mais o exame do tratamento do concepto *adecuación* nos manuais escolares debe complementarse coa aproximación ao currículo de Bacharelato que estaba en vigor no momento en que se recolleron as redaccións

que forman parte do corpus *CORTEGAL*, o chamado Bacharelato LOMCE⁵. A mención desta propiedade textual faise sempre de maneira seriada, conxuntamente coas outras:

Contidos. B.3.3. Producción de textos escritos ou orais de diferentes xéneros mostrando un correcto nivel de corrección gramatical, presentación, adecuación, coherencia, cohesión e corrección.

Criterios de avaliación. B.3.3. Producir textos escritos ou orais de diferentes xéneros mostrando un correcto nivel de corrección gramatical, presentación, adecuación, coherencia, cohesión e corrección.

Estándares de aprendizaxe. LGB3.3.1. Produce textos escritos ou orais de diferentes xéneros mostrando un correcto nivel de corrección gramatical, presentación, adecuación, coherencia, cohesión e corrección (dispoñibles en <https://www.edu.xunta.gal/portal/guiadalomce/bacharelato/materias/lca/lgl>) [última consulta: 01/02/2023].

Como se pode apreciar, no material reproducido outórgase unha grande importancia á corrección, pois esta é mencionada en dúas ocasións, distinguindo entre *corrección gramatical* e outro tipo que non se especifica. É posible que esta circunstancia derive da existencia dunha moi consistente liña de investigación, inaugurada pola tese de doutoramento de Bieito Silva Valdivia (2004) e xa na actualidade cun longo percorrido⁶, que vén detectando deficiencias na adquisición daqueles trazos gramaticais que diferencian galego e castelán, como por exemplo os verbos irregulares, os tempos verbais e a colocación dos clíticos. Os traballos formulados desde esta perspectiva conseguiron unha importante visibilidade social: son recorrentes, na prensa máis favorable ao galego, os artigos en que se menciona o non desenvolvemento dunha competencia gramatical correcta en galego por parte do alumnado⁷. Desta liña de análise despréndese o relativo fracaso do modelo de ensino de linguas vixente, ao non se conseguir o previsto no apartado terceiro do artigo

⁵ A Lei Orgánica para a Mellora da Calidade Educativa (LOMCE), en vigor entre 2013 e 2020, pode ser recuperada en <https://www.boe.es/eli/es/lo/2013/12/09/8/con>.

⁶ Nunha relación non exhaustiva, debemos mencionar Silva Valdivia (2006 e 2010); Loredó e Fernández Suárez (2005); Fernández Salgado et al. (2005); Silva et al. (2008); Loredó Gutiérrez e Silva Valdivia (2014) e máis recentemente Loredó e Silvia Valdivia (2020).

⁷ A modo de exemplo mínimo cítanse os seguintes: baseado en declaracións do propio Bieito Silva, «Os cambios morfosintácticos avisan da transformación da lingua galega» (*El Correo Gallego*, 8/12/2007, https://www.elcorreo gallego.es/hemeroteca/os_cambios_morfosintacticos_avisan_da_transformacion_da_lingua_galega-GHCG217562); canda a publicación do último estudo sobre o particular, «Un terzo do alumnado castelánfalante acaba ESO con baixas competencias en galego» (*Cultura Galega*, 8/4/2021, <http://culturagalega.gal/noticia.php?id=32417>), e «RAG: «o 35% do alumnado castelánfalante remata a ESO cunha escasa competencia en galego»» (*Novas do Eixo Atlántico*, 27/4/2021, <https://www.novasdoeixoatlantico.com/rag-o-35-do-alumnado-castelanfalante-remata-a-eso-cunha-escasa-competencia-en-galego/>).

14 da Lei de Normalización Lingüística no relativo á simetría de nivel que se debe acadar na aprendizaxe das linguas cooficiais tras o paso polo ensino obrigatorio⁸.

A combinación das dúas circunstancias que acaban de ser presentadas (unha visión da adecuación como proceso fundamentalmente asociado á selección léxica e a centralidade da corrección como medición da adquisición dos trazos sintácticos e morfolóxicos diferenciais entre galego e castelán) explica, ao noso xuízo, a escasa ou nula presenza, nos manuais deste período, de actividades que especificamente teñan como obxectivo a práctica da variación sintáctica desde a perspectiva da diversidade de rexistros. Dado que tales contidos parecen estar ausentes dos libros de texto, en que segundo vimos tende só a practicarse (e mesmo a definirse) a adecuación como proceso ligado ao compoñente léxico da lingua, unicamente o traballo en aula sobre as producións do alumnado e as correccións dos/as docentes sobre estas servirían para transmitir pautas de adecuación ligadas a outros compoñentes da lingua.

O ano 2022 representa un importante punto de modificación do paradigma de traballo escolar, tras a entrada en vigor dunha nova lei educativa, a LOMLOE⁹, e dun novo currículo de lingua galega para o Bacharelato¹⁰. Da lectura deste último documento parece desprenderse a intención de dotar de maior presenza nas aulas ao traballo de análise e construción textual; neste sentido, inclúese unha referencia explícita á «excepcional importancia» da elaboración de materiais académicos («son os que constrúen a formación científica e humanística de cada estudante nesta etapa», apartado 27.1, introdución). Para o presente traballo resulta pertinente ademais o seguinte parágrafo deste mesmo documento:

Nesta etapa prestarase especial atención a situacións con maior distancia social entre os interlocutores, que exixen usos lingüísticos máis elaborados, rexistros formais [...]. Estas destrezas serán esenciais para o desenvolvemento académico e profesional posterior [...]. As ditas secuencias didácticas atenderán ás sucesivas fases do proceso (planificación, produción, ensaio e revisión) ata chegar ao produto final (Apartado 29.2, «Obxectivos da materia», Obxectivo 3).

⁸ Velaquí a coñecida redacción deste apartado da Lei de Normalización Lingüística: «As autoridades educativas da Comunidade Autónoma garantirán que ó remate dos ciclos en que o ensino do galego é obrigatorio, os alumnos coñezan este, nos seus niveis oral e escrito, en igualdade co castelán».

⁹ A *Lei Orgánica pola que se modifica a Lei Orgánica de Educación (LOMLOE)* foi aprobada en 2020 mais non entrou en vigor en Galicia ata o curso 2022-2023. O texto da lei pode recuperarse en <https://www.boe.es/eli/es/lo/2020/12/29/3>.

¹⁰ O novo currículo, contido no Decreto 157/2022, do 15 de setembro, pode recuperarse en https://www.xunta.gal/dog/Publicados/2022/20220926/AnuncioG0655-190922-0003_gl.html.

A análise que segue pode, daquela, contribuír a dotar o profesorado de ferramentas que, convenientemente adaptadas, lle permitan traballar as cuestións concretas que aquí se analizan (os procedementos de xeneralización e marcado da implicación), no relativo á atribución dos recursos empregados polo alumnado a diferentes rexistros e á detección, por tanto, daqueles máis acaídos para o contexto académico. Por outro lado, a esencial equivalencia semántica entre estruturas de gran diversidade tipolóxica como as que se poñen en relación na presente achega pode ser obxecto dun tratamento escolar semellante ao que recibe a reiteración léxica: da mesma maneira que intentamos crear consciencia, nas persoas que aprenden lingua, de que utilizar sempre a mesma palabra dentro dun texto resulta empobrecedor, é posible tamén alertar sobre o amplo abano de recursos sintácticos con que conta a lingua para a xeneralización e o marcado da implicación, e ensinar, por tanto, as bondades da diversificación consciente de opcións estruturais.

3. APROXIMACIÓN BIBLIOGRÁFICA AOS CONCEPTOS UTILIZADOS

Deseguido realizarase unha síntese dos resultados daquelas achegas teóricas que consideramos de utilidade para a descrición do fenómeno no corpus *CORTEGAL*. As referencias consultadas sitúanse tanto na análise de producións que teñen a súa orixe na lingua dos medios, pero sobre todo na linguaxe da ciencia e da tecnoloxía e outros usos próximos (materiais divulgativos, textos escolares xerados nas aulas de diferentes disciplinas ou asociados a distintos niveis educativos¹¹), como formando parte de manuais e outros recursos de natureza práctica, cuxa finalidade é axudar os/as futuros/as redactores/as no proceso de composición das súas propias producións, sexan estas artigos de investigación, *papers* destinados a seren expostos e debatidos en reunións científicas, exames e traballos académicos etc. (Cassany, 2008). Unha subsección deste último tipo está constituída polas aproximacións que procuran axudar ás persoas que aprenden e se examinan de conceptos científicos naqueles procesos que estas/es realizan nunha lingua diferente daquela súa principal: neste ámbito sitúase a notable cantidade de achegas vinculada á *lingua en usos específicos*, cuxo desenvolvemento no mundo anglosaxón é, por razóns evidentes, vizoso¹². Como cabería esperar, a dimensión normativa aparece ligada non ás aproximacións do primeiro tipo, descritivas, senón á

¹¹ Análises de interese sobre textos semellantes aos do corpus *CORTEGAL* son Pérez Gaztelu e Zulaika Ijurko (2014) e, utilizando materiais argumentativos do mesmo grupo de idade mais vinculados á oralidade, Gimeno Martínez (2012).

¹² Unha historia do termo pode atoparse en Bolívar e Parodi (2015).

segunda categoría, en que se procura gobernar as características das producións que se xerarán.

No devandito contexto analítico teñen xurdido nas últimas décadas diversos paradigmas que se enfrontan ao marcado da presenza da persoa produtora do texto no interior da súa propia produción. Un primeiro grupo de análises, asociadas a conceptos como *metadiscurso* (Ädel, 2006; Salas Valdebenito, 2015), *atribución do coñecemento* (Hyland, 2004, 2005), *involvement* e *detachment* (Chafe e Danielewicz, 1987), *face* ou *imaxe* (Casamiglia Blancafort, 1996; Garcés Conejos e Sánchez-Macarro, 1998; Cassany, 2008) e *ethos* (Fernández Silva, 2015, p. 101), parte da visión do/a emisor/a como instancia que decide, en función de diversos factores e sempre dentro de comunidades discursivas concretas, que grao de implicación desexa manifestar a respecto dos contidos que reflicte no texto; o estudo das persoas gramaticais desde as que se emite o discurso constitúe, desde esta perspectiva, a tarefa central.

Outra visión diferente é a desenvolvida en Ciapuscio (1992, 2014), que utiliza o concepto de *desaxentivación* como paraugas para o tratamento dun heteroxéneo elenco de estruturas destinadas a eliminar as expresións de emotividade e subxectividade en textos de natureza científico-técnica ou derivados destes¹³. Son diversas as construcións que presentan, na súa opinión, tal finalidade: voz pasiva e pasivas con *se*, formas non persoais dos verbos, nominalizacións (estudadas en Muñoz, 1999), persoas gramaticais diferentes da P1 etc. Aínda que a autora insiste na condición de graduais que caracteriza os procesos de desaxentivación, é inevitable vincular a esta liña aqueles traballos en que se dá conta, mediante diversas ferramentas conceptuais, do persistente (e ata certo punto inconsciente ou non desexado) afloramento da persoa emisora; é o caso de achegas que desenvolven o concepto de *inscripción de la persona* (Venegas et al., 2013) ou *inscripción del autor* (Fernández Silva, 2015), e a denominación *emergencias del emisor* (Gimeno Martínez, 2012, p. 251). Tamén situamos neste paradigma de «obxectividade por defecto» as diversas achegas que abordan, xeralmente de maneira comparada, as notables diferenzas entre áreas de coñecemento a respecto da tolerancia no mantemento de recursos subxectivizadores (Hyland, 2005, p. 10¹⁴; Álvarez

¹³ Álvarez López (2013) utiliza o concepto *despersonalización* nun sentido semellante, trazando a súa historia e achegando unha extensa información bibliográfica.

¹⁴ De interese son, neste traballo e na páxina mencionada, os comentarios metalingüísticos feitos por investigadoras/es das ciencias («I feel a paper is stronger if we are allowed to see what was done without 'we did this' and 'we think that'»), fronte aos/ás das áreas de coñecemento humanísticas («I bring to bear on the problem my own experience»).

López, 2012; Barraión López e Lavale Ortiz, 2013; Salas Valdebenito, 2015, p. 96); así mesmo, non son infrecuentes as apelacións, en termos de gramática normativa, a unha presenza máis intensa da persoa emisora na construción do discurso, especialmente cando se trata de aprendices/as (Côrte Vitória, 2018, p. 113).

Canto ao posicionamento desde o que se realiza a presente achega, desde unha visión ecléctica tratarase de tomar aqueles aspectos de ambas que resulten de maior utilidade para a análise do corpus *CORTEGAL*. O concepto de *desaxentivación* permítenos estudar a ocorrencia da P1 nos nosos textos como excepcional, e ver, por tanto, a rendibilidade e condicións de aparición desta a respecto do resto de persoas gramaticais empregadas nas redaccións. Porén, non se tratarán diversos recursos que forman parte tamén das aproximación baseadas neste concepto, como as nominalizacións ou o emprego das formas verbais non persoais, entre outros.

Pola contra, o recoñecemento da centralidade das persoas gramaticais na expresión do posicionamento do/a emisor/a permite que aflore analiticamente un tipo de estruturas, que chamaremos no noso traballo de *xeneralización*¹⁵, caracterizadas conxuntamente por supoñeren a presentación dunha idea, punto de vista, tendencia etc., como lugar común ou opinión compartida por todos, ou por unha parte significativa, dos individuos que conforman a sociedade, ou un grupo deles. Trátase de exemplos como os reproducidos anteriormente, e mais como os que seguen:

- (12) [Esta moda pola cociña]. Na miña opinión quedará mentres *a xente* atope tempo suficiente para cociñar, pero en canto *a xente* se canse de utilizar o seu tempo libre en cociñar e apareza outra moda máis nova desaparecerá
- (13) Tamén cómpre destacarmos que ese «tráfico de comida» que se denuncia no texto é en realidade o que demanda *a poboación*, sobre todo a máis xove

Estudárase a rendibilidade dunhas e outras posibilidades construtivas, e tamén as condicións de aparición e relevancia cuantitativa daquelas estruturas en que ao valor de xeneralización se une o marcado da implicación da persoa emisora. No exame de cada unha delas e mais na comparación entre as súas circunstancias de ocorrencia quedará tamén de manifesto a posible existencia, ou non, de pautas relacionadas coa adecuación de cada opción ás características esixidas a un texto formal e cuxa avaliación é relevante.

¹⁵ O concepto é similar ao que Poudat e Loiseau (2005, p. 58), partindo de traballos anteriores de Benveniste, denominan *universalización* e utilizan para analizar algúns usos da primeira persoa en textos filosóficos, en que esta ten por función «to refer to any subject (...) [e] both autor and reader are included in this universal reference».

4. XENERALIZACIÓN E MARCADO DA IMPLICACIÓN NO CORPUS *CORTEGAL*

No marco da etnografía da comunicación e en referencia a estruturas similares ás que se analizan neste traballo, Casamiglia Blancafort (1996, p. 64) sinala que a autoridade e a credibilidade do/a emisor/a son cruciais para a súa propia lexitimación, e a posición de cada un/unha deles/as no acto comunicativo exprésase mediante as escollas lingüísticas que realiza. Aplicando esta idea aos textos de *CORTEGAL*, podemos ver que existe unha desigualdade dupla entre as persoas emisora e receptora: en primeiro lugar, á condición de persuasivos que caracteriza os materiais élle inherente tal desemeianza; o desequilibrio vese reforzado, ademais, polo feito de pertenceren os materiais a un exame de tanta transcendencia como a proba de acceso á universidade.

Agárdase, daquela, que o/a emisor/a autorice a maior parte das afirmacións formuladas mediante un contorno de opinións compartidas que lle sirva de esteo. A dupla debilidade sinalada debera promover o emprego de recursos de desaxentivación e non fomentar, por tanto, o marcado da implicación: non só se trata de expoñer opinións, senón que se está sendo avaliada/o pola forma escollida para expoñelas. Deseguido veremos ata que punto esta predición se cumpre nos materiais examinados.

4.1. A alternancia entre persoas gramaticais para a consecución de xeneralización e marcado da implicación: P1¹⁶, P2, P4 e P6 en *CORTEGAL*

Alén das persoas non ligadas aos/ás participantes no acto comunicativo, comunmente empregadas nos textos persuasivos para a exposición dos contidos previstos, as redaccións do corpus manifestan a escolla entre os recursos de visibilización proporcionados polas tres persoas gramaticais en que participa o/a emisor/a, ademais dunha presenza cualitativamente significativa da P6 con valor de xeneralización.

Segundo se expuxo anteriormente, a utilización da P1 correspóndese coa intención de deixar clara a opinión da persoa que escribe a redacción, renunciando a amparar o parecer desta nalgún tipo de tendencia conxunta e remarcando o carácter individual do seu punto de vista. Por contra, tanto P2 como P4 permiten situar as afirmacións que conteñen nun horizonte de xeneralización, ao tempo que se deixa clara a implicación da persoa emisora nas opinións, actitudes, comportamentos etc. que se transmiten.

¹⁶ Para a referencia ás persoas gramaticais empregaranse ao longo do traballo as seguintes abreviacións: P1 (primeira persoa) P2 (segunda persoa), P3 (terceira persoa) P4 (cuarta persoa ou primeira de plural), P5 (quinta persoa ou segunda de plural) e P6 (sexta persoa ou terceira de plural).

Un medidor da tendencia subxectivizadora que supón a presenza da P1 consiste en analizar a aparición desta mediante as categorías *pronome persoal* e *posesivo*. No referido ao pronome tónico e dado que este elemento resulta prescindible por causa da concordancia, a súa presenza utilízase para remarcar intensamente a subxectividade desde a que se fala. Non é infrecuente rexistrar no corpus exemplos de dupla marca subxectivizadora:

- (14) Desde o meu punto de vista, *eu* creo que a cociña cara, selecta e de pouca cantidade é unha moda pasaxeira

Localizouse un total de 174 ocorrencias de *eu* en 129 textos distintos, o que supón que (como media) en máis dun 10% das redaccións se emprega a unidade que implica unha emerxencia máis poderosa da/o emisor/a no texto¹⁷. Aparece a P1 naquelas ocorrencias en que se decide achegar a propia opinión, xunto a verbos ligados ao acto de expoñer un punto de vista persoal: así, a secuencia *eu creo que* é empregada en 51 ocasións, *eu penso que* en 38 (e unha ocorrencia de *penso eu*), *eu opino que* en cinco casos e *eu considero que* en tres.

Canto ao clítico *me*, indicador igualmente de interpretación subxectiva dos contidos, o seu número é algo inferior: 119 exemplos no total do corpus, do tipo do que segue:

- (15) Os rapaces queren ser o que se lle mostra nos medios de comunicación que son futbolistas e modelos; pero pregúntome, onde están os bombeiros, médicos,...?

O emprego dos posesivos *meu* e *miña* é tamén un indicio da inequívoca presenza da persoa emisora no texto. En 98 casos aparece o primeiro deles, case sempre ligado á estrutura *desde/dende o meu punto de vista*:

- (16) Dende *o meu punto de vista* a calquera cousa que enriqueza culturalmente é beneficioso para a sociedade

A frecuente estrutura *na miña opinión* explica case a totalidade (169) das 190 ocorrencias do posesivo feminino de primeira persoa:

- (17) Na *miña* opinión, é moi posible que este tipo de comida pase de moda

Para poder valorar cabalmente a entidade destes datos, debemos achegar as cifras das outras persoas gramaticais que implican tamén unha intención

¹⁷ O corpus achega un total de 761 exemplos de verbos en P1, polo que as ocorrencias con suxeito pronominal explícito son o 22,8% do total.

subxectivizadora e nas que converxen, segundo a nosa hipótese, xeneralización e marcado da implicación. En primeiro lugar examínase a P4, forma historicamente ligada á atenuación cortés mediante o concepto de *plural de modestia* (Freixeiro Mato, 2000, pp. 189-190; Álvarez e Xove, 2002, p. 540). A bibliografía sobre o emprego da P4 en lugar da P1¹⁸ vén detectando nesta posibilidade estrutural unha polisemia potencialmente ambigua: *nós* pode referirse unicamente a *eu* como forma atenuadora da alusión a si mesmo/a do/a redactor/a, mais tamén se refire, e trátase do seu significado primario na definición das gramáticas, a *eu* e *ti* (P1 + P2) e a *eu* e *elas/eles* (P1 + P6). Estas dúas últimas opcións non parecen posibles nos materiais do corpus *CORTEGAL*, pois requiren de condicións características da lingua oral.

O emprego da P4 para a expresión das propias ideas baixo a aparencia de puntos de vista, opinións etc. comúns supón combinar os valores de xeneralización e marcado da implicación: a subxectividade do *eu* preséntase como tendencia xeral, compartida cun grupo de persoas do que mesmo podería formar parte o/a receptor/a do texto. Os 269 exemplos localizados no corpus *CORTEGAL* do pronome *nós*¹⁹ (e dúas ocorrencias da forma composta con *outros*) sitúanse neste valor do pronome; non aparecen ocorrencias do chamado *plural de modestia*:

- (18) O consumo e a produción dependen o un do outro e, na sociedade en que vivimos, *todos nós* dependemos de ambas
- (19) Con todo, os medios de comunicación, a publicidade, etc fan incrementar este problema e é necesario darnos conta de que detrás diso, estamos *nós, as persoas*

Canto ao número de clíticos de P4, o corpus *CORTEGAL* dá como resultado a moi importante cifra de 1064 ocorrencias, todas elas situadas na combinación entre xeneralización e marcado da implicación, e sen que aparezan tampouco ocorrencias de plural de modestia:

- (20) Por iso tamén o autor deste artigo di que todos somos consumidores. Edúcanos para mercar e gañar cartos

No referido ao emprego dos posesivos *noso* e *nosa*, este é tamén máis que notable: 514 casos da forma de masculino e 784 do feminino:

- (21) Tal e como está agora mesmo o mercado e os salarios na *nosa* sociedade actual, os consumidores vémonos obrigados a apostar polas marcas brancas

¹⁸ Ciapuscio (1992, p. 197) e Casamiglia Blancafort (1996, p. 68) ofrecen tipoloxías de uso da P4 de base semántica.

¹⁹ O corpus achega un total de 3230 ocorrencias da P4, polo que nesta persoa unicamente o 8,3% dos exemplos aparece con suxeito pronominal explícito.

Parece daquela evidente que o recurso á P4 é rendible nas redaccións do corpus *CORTEGAL*. Sinalaremos finalmente que nalgunhas aproximacións, elaboradas desde unha perspectiva normativa e no contexto do castelán, considérase pouco axeitada esta posibilidade na confección de textos académicos²⁰. Para o galego non temos localizada disposición ningunha que se oriente en contra do recurso, mais tampouco non observamos a existencia de ditames favorables.

O cómputo de recursos empregados para a integración da persoa escritora nos contidos que elabora non estaría completo sen o exame do vigor cuantitativo da P2 nos textos do corpus *CORTEGAL*. Segundo veremos, non son infrecuentes secuencias deste tipo:

(22) A publicidade está en todas partes e non somos ningún se non tes un «iphone» na man e vistes unhas «adidas»

A posibilidade estrutural aparece recollida na gramática descritiva do galego, en que se alude ao uso indicando que «o locutor implica o alocutario situándose na ficción de que el é esa persoa indeterminada á que se fai referencia; deste xeito achega os feitos ao receptor da mensaxe e, xa que logo, a si mesmo como emisor» (Álvarez e Xove, 2002, p. 542), mais sen achegar pauta ningunha de natureza normativa.

Esta opción construtiva ten sido obxecto na segunda década do século XXI de diversas análises na gramática do castelán, especialmente centradas no seu emprego nos medios de comunicación²¹. Nelas destácase o carácter progresivo do que se denomina nelas *ti obxectivoador*, sinálase a súa maior frecuencia en textos orais e no español dos medios e da publicidade e detéctase a súa coexistencia a respecto con diversos elementos gramaticais que actúan como *indutores da xenericidade* ou, noutras achegas, *construtores de*

²⁰ «Es cierto que cualquier lector puede interpretar correctamente el valor generalizador de las formas de primera persona de estos fragmentos, esto es, entendiendo que los nosotros no significan “yo, el escritor, y un grupo cerrado de personas” sino “todo el mundo”. Sin embargo, es cierto también que este mecanismo de generalización resulta poco adecuado en cualquier texto de carácter formal. En otras situaciones comunicativas como, por ejemplo, una conversación coloquial, una manera cortés de generalizar es ponerse uno mismo como ejemplo, con lo que en contextos espontáneos estas estructuras resultan muy útiles. No obstante, en textos académicos, y en general en textos científico-técnicos, las normas de cortesía son diferentes: lo más adecuado es evitar que los textos señalen hacia el escritor o hacia el lector, es decir, lo conveniente es desfocalizar la atención de las personas implicadas en la comunicación a fin de enfocar con más claridad y precisión el mensaje, la información que hay que transmitir» (Montolío e Santiago, 2000, p. 153).

²¹ Hollaender Hensen (2003); Aijón e Serrano (2014); Serrano Montesinos (2013, 2021a, 2021b); Aijón Oliva (2020a).

espazos mentais, tales como o tempo verbal presente ou as construcións con *si* e con *cuando*. Desde a perspectiva de gramática normativa si se achega a consideración da P2 en funcións de xeneralización como inadecuada á marxe da informalidade, pois afirmase que «constituye un mecanismo de generalización demasiado coloquial» (Montolío e Santiago, 2000, p. 162).

Vexamos deseguido o valor cuantitativo que acada a forma no corpus *CORTEGAL* mediante o cómputo das ocorrencias de *ti* e das estruturas con clíticos de segunda persoa *te* e *che*. Localízanse 35 exemplos do pronome tónico²², xunto a 87 do pronome átono de acusativo e 68 do de dativo:

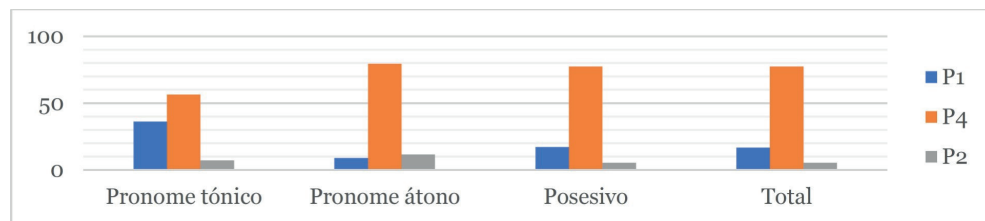
- (23) Se *ti* traballas, alguén consumirá o que *ti* produces, e precisará cartos...
- (24) Os salarios son maiores e pódeste gastar máis cartos que antes
- (25) Non é o mesmo acudir ao teu mesón de confianza (...) que sentarse nunha terraza en sillóns masaxeadores mentres tráen*che* comida á carta
- (26) Cada vez esta a crecer o número de restaurantes con estas características e a maioría moi baratos non *che* son

A rendibilidade dos posesivos *teu* e *túa* é similar: 51 ocorrencias do primeiro, 41 do segundo:

- (27) Por moi bó cociñeiro que sexas se non tes un gran material o *teu* traballo non sería tan gozado e tampouco terá tanto éxito
- (28) Por moitas gañas que teñas de ser futbolista, antes están os *teus* estudos e a *túa* formación cultural e intelectual

A Gráfica 1 ofrece a tradución porcentual e centesimal das ocorrencias de cada unha das persoas gramaticais no corpus *CORTEGAL*, tanto baixo a forma de pronome persoal como de posesivo²³:

GRÁFICA 1. Total de referencias explícitas, mediante o pronome persoal ou o posesivo, á persoa emisora do discurso



²² O número total de formas verbais que presenta o corpus en P2 é de 473, polo que nesta persoa só aparece con suxeito pronominal explícito o 7,4% dos casos.

²³ Para a elaboración das porcentaxes sumáronse as formas de posesivo masculino e feminino, así como os clíticos de acusativo e dativo da P2.

Despréndese dos datos o predominio cuantitativo da P4, que se achega ou supera os tres cuartos do total de mencións explícitas da persoa emisora tanto en números absolutos como nas categorías *pronome átono* e *posesivo*: só a subxectivización extrema que supón a mención da P1 mediante o recurso ao pronome tónico fai que este nivel de emprego se reduza. O máis usual no corpus, por tanto, é combinar xeneralización con marcado da implicación mediante o recurso a esta persoa gramatical, en tanto que a outra posibilidade, o emprego da P2, presenta un vigor non por moito máis reducido carente de significado: no total, preto do 10% de mencións explícitas se fan mediante o recurso a esta persoa gramatical, a pesar do posible ditame normativo ambiguo a respecto desta opción construtiva.

Non podemos obviar neste apartado a necesidade de aludir ao emprego, minoritario no corpus *CORTEGAL* mais non inexistente, da P6 como estratexia de xeneralización sen marcado da implicación. Trátase de ocorrencias en que a persoa vén determinada unicamente pola morfoloxía verbal e nas que, ou ben o antecedente da relación anafórica (usualmente dotado de valor totalizador) está excesivamente lonxe, ou non concorda gramaticalmente (*vid.* ao respecto o capítulo a cargo de Cidrás neste mesmo volume), ou ben é difícil determinar con exactitude cal sexa este, a pesar de que o significado da estrutura é totalmente comprensible. Falamos de exemplos como os que seguen, sobre os que cómpre indicar que en (29) e (30) se marcou con cursiva a forma comentada e mais o probable elemento anafórico a partir do cal se xustifica o emprego desta P6:

- (29) A partir destes programas xorde *xente* que se cre coñecedora da cultura de *todo o mundo* pero isto non é así e creo que cando *se decaten* de que hai outros tipos de cultura ademais da gastronómica estes programas deixarán de ser tan populares
- (30) Existen *restaurantes* nos que a cantidade de comida que *poñen* é tan reducida que casi é difícil de ver
- (31) Eso fixo que media España collera as sarténs e cacerolas e foran as cociñas para non sair de alí ata ter o seu 'solomillo con patacas cocidas e guarnición de boletus' ao punto do crítico máis cercano, é dicir, a familia, ou en casos máis extremas, os veciños. Posteriormente *estrenarían* a versión para nenos dese programa

No primeiro exemplo dos reproducidos non existe concordancia gramatical con *xente*, aínda que parece clara a condición de antecedente que presenta este elemento; en (30) a dificultade vén dada por ter o probable antecedente unha semántica inadecuada para o verbo escollido, dada a condición daquel como [-humano]. O exemplo (31) é aínda máis complexo, pois non existe elemento ningún ao que se poida atribuír o ligame anafórico, máis alá dunha

activación contextual da relación entre medios de comunicación e gastronomía que permita a presenza do verbo *estrear*, así como a dedución do seu suxeito.

Este tipo de emprego xenérico da P6 difire dos anteriormente estudados para a P4 e P2 na non implicación do/a emisor/a; é máis, podería ser asociado a unha intención de marcar a distancia entre este/a e os individuos non determinados que realizan a acción do verbo: en (29) enténdese que a persoa que redactou o texto non se considera parte do colectivo que debe decatarse do seu erro, e en (30) e (31) o afastamento ten que ver coa non condición de profesional da restauración ou dos medios de comunicación que caracteriza o/a emisor da mensaxe. Por unha circunstancia ou por outra, o emprego da P6 permite colocar a tendencia, ideoloxía, comportamento etc. da que se fala nun lugar distante do suxeito emisor. Segundo acertadamente afirma Aijón Oliva (2020b, p. 8) no seu tratamento destas estruturas no castelán, o valor inherente á persoa consiste en que os referentes indicados non son «neither you nor I».

A modo de cata que permita deducir o vigor numérico desta opción construtiva, examináronse as primeiras cinco redaccións de cada unha das 25 comisións de avaliación a partir das cales se constituíu o corpus *CORTEGAL*; nestes 125 textos foron localizados un total de 13 exemplos de P6 con valor xeneralizador como os anteriormente mencionados; de media, algo máis do 10% dos materiais presentaría un exemplo deste tipo, aínda que varios deles aparecen situados na mesma redacción, polo que o recurso, mesmo sendo minoritario, non é inédito nos textos formais que se compilaron.

4.2. A xeneralización desde o individuo: *un, cada un, cadaquén*

Freixeiro Mato (2000, p. 253) afirma que «pode tamén *un* equivaler expresivamente a ‘calquer persoa’, baixo a que se pode agochar ou que pode encubrir eufemisticamente o propio eu». Similar tratamento gramatical aparece en Álvarez e Xove (2002, p. 88). De acordo coa terminoloxía empregada na presente aproximación, entendemos que nestes usos do indefinido *un* está sempre implicado o proceso de xeneralización, pero non se produce de maneira explícita o marcado da implicación, aínda que o valor é deducible a partir do contexto. Así, en (32) parece plausible a sinonimia entre a estrutura utilizada e as opcións que se presentan en (32a), en que se marca explicitamente a concordancia entre o comportamento xeral e o escollido pola persoa redactora:

- (32) Na sociedade actual, a tendencia é valorar unha aparencia física fronte a personalidade e os gustos que *un* ten
 (32a) Na sociedade actual, a tendencia é valorar unha aparencia física fronte á personalidade e os gustos que *temos/tes*

Porén, a ausencia dun elemento lingüístico explícito que marque a implicación da persoa redactora provoca que a estrutura poida ser tamén considerada sinónima a respecto doutras en que unicamente está presente o valor xeneralizador:

- (32b) Na sociedade actual, a tendencia é valorar unha aparencia física fronte á personalidade e os gustos que *se* teñen

En calquera caso, o aspecto máis relevante no emprego da unidade é a súa moi escasa rendibilidade. Examinadas a totalidade das ocorrencias (adxectivais e pronominais, 3021) da forma, unicamente foron localizados 10 exemplos. Algo mellor numericamente é o balance que achega a fusión da individualidade coa distribución en exemplos con *cada un* (33 casos).

O reducido vigor do recurso podería poñerse en relación coa dificultade que presenta en termos de linguaxe non sexista, e actuar, por tanto, como indicio de que esta cuestión é tratada nas aulas de Secundaria. Outra hipótese que xustificaría o exíguo número de exemplos detectado consiste en mencionar como argumento a posible decadencia de *un* pronominal no rexistro informal da lingua, aínda que sería precisa unha indagación específica para corroborar que tal circunstancia se produce. De ser demostrable esta última visión, a práctica desaparición de *un* no oral (cando menos na fala das xeracións máis novas) podería poñerse en relación coa súa escaseza na formalidade académica que representan os textos do corpus *CORTEGAL*. Porén, non son mellores quantitativamente os datos que achega a forma *cadaquén*, máis axeitada en termos de linguaxe non sexista e, aínda así, igualmente escasa nos textos do corpus; das seis ocorrencias que presenta a unidade, só en tres o elemento se sitúa en posición de suxeito gramatical e con valor xeneralizador:

- (33) Nun mundo e nunha sociedade onde a información están ao alcance da maioría das persoas, vivimos practicamente desinformados. *Cadaquén* ten que buscar a información do que desexe polos seus propios medios

4.3. A voz pasiva

A utilización desta posibilidade construtiva como procedemento de consecución da obxectividade discursiva é mencionada naquelas achegas en que se realiza un inventario de opcións estruturais ligadas a esta finalidade, e así

podemos vela, entre outras, na listaxe de estratexias de desaxentivación achegada en Ciapuscio (1992, p. 201) e en Montolío e Santiago (2000, p. 155).

A presenza das estruturas pasivas tradicionais no corpus *CORTEGAL* é moi reducida. A modo de comprobación e como en apartados anteriores, realizamos unha cata nas mesmas cinco redaccións que abren cada unha das 25 comisións de avaliación: nos 125 exercicios foron localizadas un total de 15 estruturas deste tipo:

- (34) A popularidade deste fenómeno pode *ser entendida* por moitos como unha simple mostra de poder ou dominación
- (35) A cociña e a gastronomía acadaron moita popularidade, e ademais por moitos é *considerada* unha arte en si mesma

Os exemplos reproducidos dan conta dunha circunstancia que comparten sete das 15 estruturas con pasivas tradicionais inventariadas (isto é, algo menos do 50% das ocorrencias): este grupo de casos contén algún erro construtivo, xeralmente asociado á concordancia. É o caso de (36):

- (36) Estos cambios no medo á cociña «laboriosa» que se da xa dende a infancia pode levar a que *sexa adoptado* culturalmente como unha norma

Temos, pois, que a voz pasiva na súa formulación tradicional non só aparece no corpus *CORTEGAL* de maneira certamente esporádica, senón que ademais resulta ser unha construción dificultosa para as/os aprendices de lingua canto á concordancia. Ambos os indicios sosterán a nosa hipótese de que o traballo escolar con esta opción estrutural pode non ser suficiente.

Distinto é, porén, o caso das pasivas con *se*, de ocorrencia extremadamente frecuente; examinados os resultados que achega corpus *CORTEGAL* nunha busca do elemento (3338 ocorrencias do pronome *se*, das cales aproximadamente 2000 corresponden ao valor requirido), foron analizadas as 100 primeiras²⁴ e obtivéronse un total de 45 exemplos válidos de pasiva con *se*.

No traballo cos resultados obtidos ponse de manifesto o máis amplo espectro de posibilidades semánticas e sintácticas que posúe a estrutura fronte a aqueles recursos a respecto dos cales se realizaron probas de conmutación: as persoas P4 e P2 e o indefinido *un*. En 31 ocorrencias das 45 sinaladas, a substitución da pasiva con *se* por calquera outro dos procedementos analizados deu lugar a secuencias aceptables:

²⁴ Descartáronse as estruturas en que o clítico actúa como formante léxico ou pronome reflexivo, e mais tamén os exemplos xerados con *deberse (a)*, *tratarse (de)* ou *dar(se)*. Estes últimos non forman parte, ao noso xuízo, de estruturas en que estean implicadas a desaxentivación ou a xeneralización, pois non é posible neles a conmutación pola voz activa correspondente ou pola pasiva tradicional.

- (37) A gastronomía non é só algo que *se come*, tamén é visual
 (37a) A gastronomía non é só algo que *comemos/comes/un come*, tamén é visual

As outras 14 presentaron diversos problemas de substitución, que se exemplifican en (38-40):

- (38) Cada vez máis xente viaxa habitualmente [e] podemos observar numerosos anuncios por medios nos que *se ofertan* billetes de avión, barco etc. a bós prezos
 (38a) Cada vez máis xente viaxa habitualmente [e] podemos observar numerosos anuncios por medios nos que **ofertamos/*ofertas/*un oferta* billetes de avión, barco, etc., a bos prezos
 (39) Hai xente que o fai [asumir novas modas gastronómicas] por un motivo serio ou por denunciar a forma en que *se trata* ós animais destinados a acabar nos supermercados.
 (39a) Hai xente que o fai [asumir novas modas gastronómicas] por un motivo serio ou por denunciar a forma en que *tratamos/*tratas/*un trata* os animais destinados a acabar nos supermercados
 (40) As empresas tamén producen máis do necesario (...) e finalmente moitos destes produtos son tirados ao lixo. En este proceso *prodúcese* unha perda enorme de cartos innecesaria
 (40a) As empresas tamén producen máis do necesario (...) e finalmente moitos destes produtos son tirados ao lixo. Neste proceso *producimos/produces/*un produce* unha perda enorme de cartos innecesaria

As estruturas modificadas de (38a) resultan agramaticais porque en (38) o emprego da pasiva con *se* permite ocultar a/o axente, mais non ten como función a de xeneralizar no sentido de indicar unha acción, ideoloxía, actividade etc., da que esta/e pode participar: a venda de billetes de avión lígase a un colectivo profesional concreto, alleo á persoa emisora. En (39a) si é viable a substitución da pasiva por P4, pois, aínda non sendo a/o axente real da acción, unha persoa pode asumir a súa responsabilidade no comportamento social sinalado e utilizar, por tanto, un recurso que supón a un tempo xeneralización e marcado da implicación. O mesmo ocorre en (40a) e aínda en maior medida, dado que o emprego da P2 é posible, mais non o do indefinido *un*, o cal, queda descartado polo seu significado de xeneralización desde o individuo.

As conmutacións permiten comprobar, segundo se indicou, o maior espectro semántico cuberto pola pasiva con *se*: a misión fundamental da estrutura é o ocultamento da persoa axente e, aínda que na maior parte dos casos esta circunstancia provoca que a opción construtiva sexa unha candidata ideal para a expresión da xeneralización sen marcado da implicación, as súas posibilidades de ocorrencia son certamente maiores.

4.4. Frases nominais e indefinidos

Estudaranse deseguido aquelas estruturas cuxo valor de xeneralización na descrición de opinións, costumes, criterios etc., a respecto dos cales a/o alumna/o achega o propio punto de vista, vén dado polo significado de determinadas unidades substantivais, ou máis ben das frases nominais en que estas se insiren. Parece pertinente achegar unha descrición do comportamento destas unidades no corpus *CORTEGAL*, dado que o seu emprego se converte nunha posibilidade estrutural alternante a respecto dos recursos estudados en §4.1, §4.2 e §4.3; neste sentido poden ser interpretadas as conmutacións que seguen, realizadas sobre exemplos que pertencen ao corpus:

- (41) *Moita xente* ve agora a comida como unha forma de pracer máis que de nutrición
- (41a) *Vese/Ves/Vemos/Un ve* agora a comida como unha fonte de pracer máis que de nutrición
- (42) Non é xusto valorar o estado da sociedade tendo en conta o que *as persoas* soñan na súa infancia
- (42a) Non é xusto valorar o estado da sociedade tendo en conta o que *se soña/soñamos/soñas/un soña* na *nosa/túa* infancia
- (43) Os críticos gastronómicos naceron co movemento e *a sociedade* os aceptou sen miramentos
- (43a) Os críticos gastronómicos naceron co movemento e *aceptáronse/aceptámoslos/aceptáchelos/un aceptounos* sen miramentos
- (44) Tamén cómpre destacarmos que ese «tráfico de comida» que se denuncia no texto é en realidade o que demanda *a poboación*
- (44a) Tamén cómpre destacarmos que ese «tráfico de comida» que se denuncia no texto é en realidade o que *se demanda/demandamos/demandas/un demanda*
- (45) Da produción encárganse unhos poucos durante un período de tempo pero *todo o mundo* pode consumir cando e onde queira
- (45a) Da produción encárganse uns poucos durante un período de tempo pero *consómese/consumimos/consomes/un consome* cando e onde queira
- (46) Pola contra, está a gran dependencia que *os seres humanos* temos ante o consumo e a produción
- (46a) Pola contra, está a gran dependencia que *se ten/temos/tes/un ten* ante o consumo e a produción
- (47) En resumo, creo que *todos* deberíamos valorar máis outras cousas non tan materialistas
- (47a) En resumo, creo que *se deberían/deberíamos/deberas/un debería* valorar máis outras cousas non tan materialistas

Ao optar por un elemento nominal para a expresión do valor de xeneralización, a/o emisor/a ten en conta dúas circunstancias vantaxosas a respecto das opcións anteriormente estudadas, alén da semántica intrínseca da forma e das

súas características morfolóxicas; en primeiro lugar, a introdución de elementos modificativos na estrutura nominal permite determinar o tamaño do grupo de referencia e expresar, ben unha xeneralización plena, ben unha xeneralización referida a un segmento da sociedade unicamente; en segundo termo, a persoa emisora pode provocar, ou non, o marcado da implicación, mediante o emprego da concordancia gramatical en P4 en lugar de en P3 ou P6.

4.4.1. Xente

O substantivo *xente*, común con este valor no rexistro coloquial da lingua e de significado colectivo (e polo tanto, previsiblemente non pluralizable)²⁵, aparece como o recurso máis rendible, en compañía do artigo determinado, á hora de situar un elemento nominal con función xeneralizadora. O corpus *CORTEGAL* permite localizar un total de 216 exemplos do tipo de (48):

(48) Grazas á produción, *a xente* pode consumir, é dicir comprar

O verbo que acompaña a frase nominal está case sempre en presente, de indicativo ou subxuntivo, e xeralmente en P3; porén, localizáronse dúas ocorrencias de concordancia con P6²⁶:

(49) Sen consumo e sen produción *a xente* non *terían* diñeiro para vivir

Segundo se indicou anteriormente, o emprego desta unidade permite o establecemento de subconxuntos, fundamentalmente cuantitativos mais tamén cualitativos, a respecto da totalidade ilimitada que supón a utilización da forma núa *xente*. *A maioría da xente, a meirande parte da xente, non toda a xente* son expresións deste tipo; obsérvase a concordancia con P4²⁷:

(50) Moitas veces os adolescentes e quizáis *a maioría da xente* empregamos a táctica de se eso non me me funciona con un chamo ao outro

Canto á creación de subconxuntos de base non cuantitativa, exemplifícanse mediante ocorrencias do tipo de (51):

(51) O máis importante é o que pensa *a xente normal*, a poboación

²⁵ O plural *xentes* localízase no corpus, mais nunca en estruturas deste tipo.

²⁶ Dous exemplos máis sitúanse de a cabalo entre os examinados neste apartado e os analizados en liñas anteriores como de P6 con valor xeneralizador, así en: *Agora somos máis descuidados co diñeiro, non nos importa gastar e antes a xente prefería aforrar, xa que viviron anos moi duros.*

²⁷ A mesma dupla posibilidade de concordancia detéctase no portugués europeo (Rubio, 2014) e brasileiro (Scherre et al., 2018).

Sinalaremos tamén a aparición no corpus *CORTEGAL* dunha estratexia que serve para marcar a distancia entre o grupo sinalado e a persoa emisora, isto é, para a expresión da non implicación. Trátase do emprego das frases *esta xente* e *esa xente*, en que o demostrativo ten como función precisamente a de servir como marca de que a persoa emisora non pertence ao grupo que se menciona, isto é, de non implicación:

(52) Se a produción diminúe *esa xente* terá que buscar o que quere consumir noutro lugar

4.4.2. Persoas, seres humanos, homes, cidadáns e individuos (*en singular e plural*)

O emprego destas unidades substantivais en estruturas do tipo das que nos interesan presenta como nexa de unión a aparición en plural do elemento nominal, aínda que (segundo veremos) o emprego dun singular con valor xenérico é tamén posible para algunhas das unidades examinadas. O colectivo sobre o que se establece unha afirmación global enténdese, daquela, como constituído por unidades menores diferenciadas, polo que xeralmente o verbo aparece concordando en P6 co substantivo en plural.

Dentro desta agrupación de elementos nominais, o sintagma *as persoas* é con diferenza o máis rendible, pois ocorre en 108 ocasións co valor perseguido, como en (53):

(53) Nun futuro non moi lonxano *as persoas* xa non traballarán en fábricas

Tamén a frase é matizable cuantitativamente (*todas as persoas, a maioría das persoas*) e cualitativamente, como en (54),

(54) Normalmente *as persoas que son viaxeiras* interésanse moito pola cultura de cada país, e a cultura tamén é a gastronomía

O emprego deste substantivo ten como peculiaridade principal o feito de se localizaren no corpus *CORTEGAL* un total de 22 exemplos (algo máis do 20%) en que a concordancia se fai con P4, como en (55):

(55) Hoxe en día, *a maioría das persoas consumimos* máis do que producimos

E, ao igual que no caso anterior, existe a posibilidade de marcar a distancia entre a persoa emisora e o grupo do que se está falando mediante o emprego dos demostrativos na función de determinación; así en (56):

(56) Debido a globalización rexéitase o propio (...). Esta actitude é común porque *esas persoas* non sufriron o bombardeo continuo audiovisual de ese produto

As outras posibilidades construtivas ligadas ao emprego de substantivos en plural e mencionadas no título do presente subapartado presentan unha rendibilidade considerablemente menor, así como unha dupla posibilidade de canto ao número gramatical; en efecto, as características semánticas destas unidades permiten tanto contemplar o colectivo de maneira global (empregando, por tanto, o plural: *os seres humanos, os homes, os cidadáns, os individuos*) como utilizar unha forma de singular en que o exemplar-modelo serve para a xeneralización (*o ser humano, o home, o cidadán, o individuo*). Cómpre anotar, igualmente, o feito de que cando menos dúas das unidades da serie (a segunda e a terceira) poden presentar dificultades en termos de linguaxe non sexista, polo que coñecer a rendibilidade que as caracteriza no corpus *CORTEGAL* é importante tamén como mostra do posible avance de alternativas didácticas que se ocupen desta cuestión.

Dos tres sintagmas sometidos a medición, é *o ser humano/os seres humanos* o que ocorre con algo máis de entidade: 31 exemplos en singular, nove en plural:

- (57) *O ser humano* é capaz de consumir canto produce
 (58) *Os seres humanos* necesitan consumir

A presenza no corpus *CORTEGAL* dos sintagmas *o home/os homes* co seu tradicional valor de referencia á totalidade dos seres humanos é moi reducida: só se localizan sete exemplos en singular e un en plural:

- (59) Eu creo que gran parte do noso interese pola cociña ten a súa causa nas cores dos alimentos, xa que está demostrado cientificamente que *o home* se sente inclinado polas cores

En dous casos máis, a consecución do valor xenérico co substantivo que marca de xeito explícito o xénero masculino deriva da inclusión no sintagma da forma de xénero feminino:

- (60) Dende o principio da historia *os homes e mulleres da nosa sociedade* dedicáronse a produción de alimentos e útiles básicos para a supervivencia

Similares posibilidades de ocorrencia en estruturas xeneralizadoras détécense para os substantivos *ciudadán e individuo*, que poden ser usados para esta finalidade tanto en singular como en plural. A rendibilidade de ambos é, de novo, moi reducida; para o primeiro, 14 exemplos (catro en singular, 10 en plural) presentan o significado requirido en compañía do artigo definido

ou outros elementos determinativos, e admiten tanto a concordancia verbal en P4 ou P6:

- (61) *Todos os cidadáns* protestamos pola actitude dos políticos
- (62) Hai persoas que están a pasar fame ou que non teñen un lugar no que vivir mentres que *moitos cidadáns* malgastan o seu diñeiro mercando por exemplo un iPhone

Canto a *individuo*, 11 ocorrencias (sete en singular, catro en plural) presentan o significado que nos interesa; a concordancia con esta unidade nunca se fai mediante a P4:

- (63) Cada *individuo* debe valorar se considera máis importante un móvil ou as mensaxes que lle envía o seu namorado expresando todo o que sinte por ela

4.4.3. Sociedade, poboación, cidadanía, mundo

Outro recurso máis, detectado tamén para a expresión de valores xenéricos no corpus *CORTEGAL* consiste no emprego dos substantivos abstractos *sociedade*, *poboación* e *ciudadanía*, reinterpretados neste caso como concretos e colectivos. *A sociedade* como suxeito plural de actitudes, pensamentos, puntos de vista ou mesmo accións aparece en 124 ocorrencias. Trátase de secuencias con valor xeral, como en (64), ou restrinxido a un subgrupo de individuos, como en (65) e (66):

- (64) Isto fai que *a sociedade* queira sempre os artigos máis novos
- (65) Fronte a outros tempos nos que *a sociedade galega* buscaba o propio abastecemento, na actualidade o que máis se prima e a produción para outros
- (66) Todos coñecemos a Inditex e *case toda a sociedade* merca a súa roupa alí

Pola súa parte, para *a poboación* localízanse 44 ocorrencias con este significado. Cómpre destacar o predominio das estruturas de significado partitivo: *a maioría da poboación*, *unha parte moi pequena da poboación* e similares achegan 25 exemplos (o 56,8% do total):

- (67) *A gran maioría da poboación*, por non dicir o 99 % de a poboación, vive dentro dos márxenes do consumismo

No caso de *ciudadanía* só se localizou un exemplo con este valor:

- (68) Comparto a súa idea [da autora] de que *a cidadanía* debería ser capaz de valorar máis o traballo de científicos, artistas e escritores en concreto de Galicia

Debemos chamar tamén a atención sobre a ocorrencia da estrutura *todo o mundo* con valor xeneralizador. A diferenza das anteriores, a expresión

xorde do rexistro coloquial da lingua mais, a semellanza destas, o seu emprego require dunha reinterpretaoión da forma base *mundo* como substantivo colectivo, sempre co apoio do indefinido e o artigo. Localizáronse un total de 27 exemplos do significado perseguido:

(69) Actualmente, *todo o mundo* debe producir para a súa supervivencia

Tamén neste caso é posible amosar a implicación da persoa produtora da mensaxe no grupo xeral establecido pola unidade mediante o emprego da P4:

(70) *Case todo o mundo*, sobretudo os estudantes, *sabemos* o que é chegar ao máis alto da sociedade na cultura en calquer rama

4.4.4. Todos, moitos, algúns, poucos, alguén, ninguén

Achéganse deseguido os resultados da análise das ocorrencias do pronome *todos*, que aparece con este valor totalizador en 172 ocasións no corpus *CORTEGAL*. Dous aspectos resultan de interese: en primeiro lugar, no balance entre a concordancia en P4 ou P6 destaca o amplo predominio da primeira (130 ocorrencias, máis do 75% do total). O recurso emprégase maioritariamente, por tanto, para marcar a implicación e pertenza da persoa emisora ao grupo amplo establecido mediante o indefinido, e así ocorre en (71); unha neutralidade a respecto da integración do grupo xeral, como en (72), é tamén posible, aínda que menos frecuente:

(71) Como ben di o autor, *todos* somos consumidores

(72) *Todos* se merecen algunha satisfacción neste mundo

En segundo termo, sinalamos a case inexistencia da estratexia atenuadora, recomendada desde a perspectiva da linguaxe non sexista, consistente no emprego da estrutura *todos e todas*, que ocorre unicamente en tres ocasións (*todas e todos* non se rexistra):

(73) O mundo da gastronomía está en expansión e *todos e todas* queren participar del

Da mesma maneira que os substantivos permiten a expresión de valores partitivos e a delimitación de grupos menores, tamén os pronomes presentan tal posibilidade: localízanse tres exemplos de *moitos* e tres de *poucos* (deles, dous de *uns poucos*) e mais un de *algúns* con ese valor, polo que a rendibilidade destes recursos é moi reducida; repárese no emprego da persoa gramatical para marcar ben implicación ben distancia:

(74) E *moitos* podemos chegar a esta conclusión, pero *poucos* poden levala á práctica

Porén, como elementos pronominais que permiten a expresión de xeneralizacións parciais, son as formas *alguén* e sobre todo *ninguén* aquelas para as que detectamos un número algo significativo de ocorrencias en estruturas de xeneralización. En concreto, o corpus *CORTEGAL* ofrece 16 da primeira forma e 43 da segunda:

(75) Se *alguén* ten 3 mobiles é que vive mellor que *alguén* que ten 2 mobiles

(76) Nos anos 20 produciuse o auxe do consumo, *ninguén* tiña nada e tiñan que mercalo todo novo

4.4.5. *Algunhas exploracións estatísticas de interese*

Neste apartado final de recapitulación ofrecerase unha análise cuantitativa que procurará dar resposta a diversas cuestións relacionadas coa rendibilidade, condicións de adquisición e dificultades específicas no funcionamento das opcións nominais para a xeneralización cuxas condicións de emprego no corpus *CORTEGAL* foron analizadas nos apartados §4.4.1 a §4.4.4.

Obsérvase que as fórmulas máis empregadas na xeneralización mediante elementos nominais ou pronominais son *a xente* (26,05% do total de ocorrencias), *todos* (20,74%) e, nun segundo nivel canto ao seu valor cuantitativo, *a sociedade* (14,95%) e *as persoas* (13,02%). No seguinte chanzo sitúase *a poboación* (5,30%), *o ser humano/os seres humanos* (4,82%) e, como fórmula negativa, *ninguén* (5,18%); aínda por baixo do 5% de ocorrencias, destaca a presenza da fórmula *todo o mundo* (3,25%) e, finalmente, aparecen con menos dun 2% dos casos as restantes posibilidades construtivas: estúdanse 16 recursos e a metade concentra por riba do 90% dos exemplos.

Das oito opcións máis rendibles, a terceira, cuarta, quinta e sexta, isto é, as estruturas *a sociedade*, *as persoas* e *a poboación* e *o ser humano/os seres humanos*, son, en termos de *adecuación*, propias de rexistros formais da lingua, en tanto que as demais se comparten coa linguaxe coloquial. A entidade cuantitativa destas estruturas frecuentemente activadas e propias da formalidade é notablemente inferior a respecto da correspondente ás outras catro: o 55,02% dos exemplos constrúese con opcións etiquetables como coloquiais (entre elas, e aínda que cun reducido vigor numérico, *todo o mundo*); o 38,09% dos casos lígase a estruturas procedentes de rexistros formais.

No outro extremo do espectro cuantitativo, as fórmulas *o individuo/os individuos*, *a poboación* e a residual *a cidadanía*, todas de evidente adquisición mediante o contacto ou a aprendizaxe de variedades non espontáneas,

aparecen en menos dun 10% das secuencias que analizamos. A linguaxe formal ofrece, daquela, un máis amplo abano de recursos construtivos, mais o vigor das opcións comúns co coloquio, aquelas que o alumnado redactor dos materiais podería empregar para idéntica función en circunstancias de menor esixencia comunicativa, resulta ser máis que notable.

Recapítularase deseguido a análise das formas a respecto da concordancia verbal. Os comportamentos máis claros son os de *a xente* (minoritaria con P4) e *todos* (cun 75% de P4): a primeira estrutura marca distancia, a segunda maioritariamente serve para marcar implicación. *As persoas*, no esencial similar a *a xente*, aparece porén máis proclive ao emprego da P4 (20% dos exemplos); agás *os individuos*, as demais opcións admiten tamén, aínda que minoritariamente, esta concordancia.

Por último, un aspecto que consideramos de interese é o relativo á entidade daquelas alternativas máis problemáticas en termos de linguaxe non sexista: *o home/os homes*, *o cidadán/os cidadáns* e a serie de indefinidos *todos*, *moitos*, *poucos* e *algúns*. Canto ao primeiro dos elementos, o tradicionalmente empregado (e ás veces reivindicado) como expresión da neutralización de xénero, subliñouse na análise o seu mínimo valor cuantitativo (unicamente oito ocorrencias); algo menos residual é a presenza das dúas seguintes alternativas (pero ningunha alcanza o 2% de exemplos) e maior o peso da serie de indefinidos, especialmente da forma *todos*. En conxunto, as posibilidades que presentan dificultades neste aspecto resultan activadas en pouco máis dun 25% dos casos. Con todo, talvez sexa máis interesante destacar a case nula presenza dos recursos que foron ideados desde a intervención normativa para evitar unha masculinización inaxeitada: *homes e mulleres / mulleres e homes/todos e todas* non son alternativas máis que excepcionais, *todas e todos* non aparece e a posibilidade de marcar dobremente o xénero en *ciudadán* tampouco se activa.

5. SÍNTESE DE RESULTADOS E APLICACIÓN PEDAGÓXICA ARREDOR DO CONCEPTO ADECUACIÓN

Tras o exame das condicións de ocorrencia dos recursos presentes no corpus *CORTEGAL* e ligados aos procesos de xeneralización e marcado da implicación, observouse como son posibles diversas opcións estruturais: emerxencia da persoa emisora mediante P1 e expresión da implicación desta en postulados xerais mediante P4 e P2; xeneralización mediante P6 sen un antecedente claro anterior, mais deducible pola información contextual presente; estruturas formuladas na voz pasiva, tanto (minoritariamente) na súa

versión tradicional como, de maneira máis común, en pasiva con *se*, adaptadas para a expresión da xeneralización mais cun significado amplo, ligado á indeterminación do axente; moi residual emprego do pronome *un* e das formas reforzadas con distributivo *cada un* e *cadaquén*; e, finalmente, utilización de diversos elementos nominais e do pronome *todos* para crear estruturas con valor de xeneralización nalgunhas das cales é posible a indicación da implicación da persoa emisora mediante a selección do tempo verbal axeitado.

A análise achegou información tamén acerca tanto do auxe como da decadencia dalgunhas posibilidades construtivas: o indefinido *un* case non atopa espazo expresivo, mentres as opcións que supoñen o marcado da implicación do/a emisor/a, ligadas todas elas ao emprego das P4 e P2, parecen avanzar mesmo cando a responsabilidade do valor xeneralizador recae no espazo nominal (*as persoas pensamos* e similares). A pasiva tradicional resulta ser un recurso que presenta máis inconvenientes do que vantaxes, debido á pouca pericia dos/as aprendices de lingua no establecemento da relación de concordancia precisa para a localización do antecedente.

A aproximación ao concepto de *adecuación textual* que se ofrece nos primeiros apartados do presente traballo baséase en analizar o proceso como unha especie de etiquetaxe de formas e recursos, en función da cal algúns son considerados, nun momento dado da historia e para un grupo humano concreto, como aptos unicamente no rexistro formal da lingua, mentres que outros se cualifican como coloquialismos e, por tanto, só poderían ser utilizados na produción de textos asociados ao ámbito do non formal. Un terceiro tipo debería ser presentado como capaz de navegar entre rexistros ou, máis en concreto, á maneira de posibilidades construtivas que, aínda comezando a súa andaina na oralidade, xa foron consensuadas pola comunidade como acaídas tamén en contextos formais.

Para o caso concreto das estruturas que nos ocupan, o traballo de establecer a que grupo pertence cada unha delas está só parcialmente completado e, ata onde nós podemos revisar, non se concreta en formulación escolar ningunha ligada ao formato libro de texto. O emprego xeneralizador de P4 e P2 conta, na gramática normativa do castelán, con senllos ditames que o desaconsellan, mais estes non existen para o galego. Ambas as linguas comparten a conceptualización gramatical da P4 como recurso válido para a desaxentivación (o chamado *plural de modestia*), mais non se rexistran exemplos de tal uso no corpus *CORTEGAL*: de ser pertinente esta posibilidade para a gramática do galego formal, debe ser ensinada escolarmente de

xeito explícito, pois non existe na variedade coloquial da lingua. A utilización da P6 sen antecedente claro ou expreso e cun significado próximo ao que se consegue mediante a voz pasiva moi probablemente sería corrixida pola maior parte das/os docentes, aínda que non se obtivo información ningunha acerca do tipo de traballo escolar realizado con estruturas de tan elevada frecuencia no rexistro coloquial da lingua.

Por outra banda, o moi escaso emprego da voz pasiva tradicional correlaciona cunha certa impericia no seu manexo, impericia que podería requirir unha maior atención ao fomento deste recurso nas aulas. O indefinido *un* si coñece unha utilización ortodoxa desde unha perspectiva gramatical, mais cunha mínima rendibilidade: cremos que debera aflorar o debate acerca da súa procedencia ou improcedencia no rexistro formal da lingua, especialmente tendo en conta o case seguro retroceso da construción na oralidade, a súa probable asociación a variedades diastráticas propias de adultos/as maiores e as dificultades que o recurso podería supoñer en termos de linguaxe non sexista.

Canto ás secuencias estudadas no último subapartado, en que o valor xeneralizador provén da frase nominal que funciona como suxeito, na nosa opinión deberan ser tamén obxecto de traballo escolar, especialmente canto á pertinencia dos distintos substantivos no rexistro formal da lingua, á validez de cada un deles en contextos non coloquiais, a posibilidade da concordanza verbal con P4 e, por último, no que ten que ver coa formulación de expresións non sexistas. O traballo na aula con estas estruturas debera facer explícita a máis que probable valoración do grupo nominal *todo o mundo* como improcedente no rexistro formal, así como a escasamente aconsellable opción pola fórmula *a xente*, cando menos como recurso único e, en moitas ocasións, reiterado.

Os resultados da análise son compatibles coa hipótese, formulada no inicio de §4, segundo a cal na construción das redaccións que forman parte do corpus *CORTEGAL* acadan unha grande importancia as circunstancias que explican o modo de produción dos textos; así, o nulo control previo do tema, o estado mental da persoa que redacta, a desigualdade de roles entre emisor/a e receptor/a propios da linguaxe persuasiva, unidos a un talvez insuficiente traballo previo sobre os recursos que cómpre activar, explican, ao noso xuízo, a notable converxencia entre as posibilidades construtivas activadas no corpus e as características do rexistro coloquial da lingua, e tamén a menor habilidade dos/as aprendices/as no manexo daquelas construcións de moi escasa ou nula aparición no coloquio.

De cara ao traballo nas aulas de lingua galega, a evidencia de que a adecuación textual non está referida unicamente á selección léxica deberá situar a alternancia de estruturas entre rexistros no lugar que lle corresponde para un adecuado manexo da variedade formal, sempre partindo, por suposto, do carácter móbil, e pertinente só en determinado momento da historia, das clasificacións que as/os docentes e os manuais achegan ás/aos aprendices.

REFERENCIAS BIBLIOGRÁFICAS

- ÄDEL, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins. <https://doi.org/10.1075/scl.24>
- ALJÓN OLIVA, M. A. (2020a). It can be us or you. The desubjectification of viewpoint through person choice in Spanish oral and written media discourse. *Journal of Pragmatics*, 163, 4-17. <https://doi.org/10.1016/j.pragma.2020.04.009>
- ALJÓN OLIVA, M. A. (2020b). Talking about 'others': referential readings and pragmatic functions of non-phoric plural third persons in Spanish media discourse. *Lingua. International Review of Media Discourse*, 243, 1-18. <https://doi.org/10.1016/j.lingua.2020.102906>
- ALJÓN OLIVA, M. A. e SERRANO, M. J. (2014). Moving towards the realm of the other: second-person objectivization in Spanish media discourse. *Language Sciences*, 45, 173-188. <https://doi.org/10.1016/j.langsci.2014.07.001>
- ÁLVAREZ, R. e XOVE, X. (2002). *Gramática da lingua galega*. Galaxia.
- ÁLVAREZ DE LA GRANJA, M. (2018). Corpus de textos de estudantes galegos (CORTEGAL). Aspectos metodolóxicos. En M. Díaz, G. Vaamonde, A. Varela, M. C. Cabeza, J. M. García-Miguel e F. Ramallo (Eds.), *Actas do XIII Congreso Internacional de Lingüística Xeral* (pp. 55-62). Universidade de Vigo. <http://cilx2018.uvigo.gal/actas/resumos/655842.html>
- ÁLVAREZ DE LA GRANJA, M. (2020a). O infinitivo flexionado no Corpus de textos galegos escritos por estudantes no ámbito académico (CORTEGAL). *Madrygal. Revista de estudos gallegos*, 23, 15-38. <https://doi.org/10.5209/madr.73064>
- ÁLVAREZ DE LA GRANJA, M. (2020b). Análise da conxugación verbal no Corpus de textos galegos escritos por estudantes no ámbito académico (CORTEGAL). En L. de Castro Moutinho, R. L. Coimbra e A. Gómez Bautista (Coords.), *Línguas Minoritárias e Variação Lingüística* (pp. 150-173). Universidade de Aveiro. <https://doi.org/10.34624/rj68-vz44>
- ÁLVAREZ LÓPEZ, M. F. (2012). *¿Presento, presentamos o se presenta?* La autoría en el español académico escrito de CC. Económicas y Empresariales. En A. Van Hooft (Coord.), *El español de las profesiones: artículos seleccionados del IV Congreso Internacional de Español para Fines Específicos (CIEFE)*. Amsterdam, noviembre de 2011 (pp. 27-39). Ministerio de Educación, Cultura y Deporte, Subdirección General de Documentación y Publicaciones. https://cvc.cervantes.es/ensenanza/biblioteca_ele/ciefe/pdf/04/cvc_ciefe_04_0006.pdf

- ÁLVAREZ LÓPEZ, M. F. (2013). *La despersonalización en el discurso académico escrito* [Tese de doutoramento, Universidade de Alcalá]. TESEO. <https://www.educacion.gob.es/teseo/mostrarRef.do?ref=349713>
- BARRAJÓN LÓPEZ, E. e LAVALE ORTIZ, R. M. (2013). Los recursos de impersonalidad en el lenguaje jurídico y en el lenguaje médico. *Español Actual*, 99, 25-50.
- BOLÍVAR, A. e PARODI, G. (1995). Academic and Professional Discourse. En M. Lacorte (Ed.), *The Routledge Handbook of Hispanic Applied Linguistics* (pp. 459-476). Routledge.
- CASAMIGLIA BLANCAFORT, H. (1996). Multifaceted Dimensions of Self Reference. *Links & Letters*, 3, 61-76. <https://raco.cat/index.php/LinksLetters/article/view/49846>
- CASSANY, D. (1995). *La cocina de la escritura*. Anagrama.
- CASSANY, D. (2008). Metodología para trabajar con géneros discursivos. En P. Salaburu Etxeberria e I. Ugarteburu Gastañares (Eds.), *Espezialitate hizkerak eta terminologia III: espezialitate hizkeren didaktika eta komunikazioa. Jornadas sobre lenguajes de especialidad y terminología; Leioa, Biscaia, 1 febrer de 2008* (pp. 9-24). Universidad del País Vasco.
- CHAFE, W. e DANIELEWICK, J. (1987). Properties of Spoken and Written Language. En R. Horowitz e S. J. Samuels (Eds.), *Comprehending Oral and Written Language* (pp. 83-113). Academic Press.
- CIAPUSCIO, G. E. (1992). Impersonalidad y desagentivación en la divulgación científica. *LEA. Lingüística Española Actual*, 14(2), 183-206.
- CIAPUSCIO, G. E. (2014). Mecanismos de divulgación del conocimiento. En E. Montolío (Ed.), *Manual de escritura académica y profesional. II. Estrategias discursivas* (pp. 285-318). Ariel.
- CÔRTE VITÓRIA, I. (2018). *La escritura académica en la formación universitaria*. Narcea.
- FERNÁNDEZ SALGADO, X. A., LOREDO GUTIÉRREZ, X. e SUÁREZ FERNÁNDEZ, I. (2005). Análise da expresión escrita en galego e castelán ao final do ensino obrigatorio resultados dun estudo realizado na comarca de Santiago de Compostela. *Cadernos de lingua*, 27, 109-126. <https://doi.org/10.32766/cdl.27.58>
- FERNÁNDEZ SILVA, C. (2015). *La inscripción del autor en la escritura académica. Análisis de monografías escritas por estudiantes nativos y no nativos de español* [Tese de doutoramento inédita]. Universidad Antonio de Nebrija.
- FERNÁNDEZ SMITH, G. e ESCORIZA MORERA, L. (2004). Variación léxica y texto. Análisis cualitativo del grado de formalidad en contexto con AQUAD. *Pragmalingüística*, 12, 73-90. <https://doi.org/10.25267/Pragmalinguistica.2017.i25>
- FREIXEIRO MATO, X. R. (2000). *Gramática da lingua galega. II. Morfosintaxe*. A Nosa Terra.
- GARCÉS-CONEJOS, P. e SÁNCHEZ-MACARRO, A. (1998). Scientific Discourse as Interaction. En A. Sánchez-Macarro e R. Carter (Eds.), *Linguistic Choice across Genres* (pp. 173-190). John Benjamins. <https://doi.org/10.1075/cilt.158.13gar>

- GIMENO MARTÍNEZ, M. (2012). *Usos argumentativos orales en estudiantes adolescentes* [Tese de doutoramento, Universitat de València]. TESEO. <https://www.educacion.gob.es/teseo/mostrarRef.do?ref=975771>
- GÓMEZ CAPUZ, J. (2006). Evaluación de las propiedades textuales en enseñanza secundaria: la adecuación en la producción de textos escritos. *Tonos digital. Revista de estudios filológicos*, 11. <https://www.um.es/tonosdigital/znum11/estudios/8-correccion.htm>
- GONZÁLEZ AVIÓN, I. e ÍNSUA LÓPEZ, E. X. (2016). *Lingua galega e literatura 1º Bacharelato*. Baía Edicións.
- HYLAND, K. (2004). Disciplinary Interactions: metadiscourse in L2 Postgraduate Writing. *Journal of Second Language Writing*, 13, 133-151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- HYLAND, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-191. <https://doi.org/10.1177/1461445605050365>
- HOLLAENDER JENSEN, M. (2003). Expresiones impersonales y espacios mentales. *Foro Hispánico. Revista hispánica de Flandes y Holanda*, 23, 83-93. https://doi.org/10.1163/9789004334403_008
- IGLESIAS ÁLVAREZ, A., MOO PEDROSA, X. M., PENA SÁNCHEZ, X. R., RODRÍGUEZ RODRÍGUEZ, A. e ROSALES CERELJO, M. (2014). *Lingua e literatura 1º Bacharelato*. Vigo: Xerais.
- LOREDO, X. e SILVA VALDIVIA, B. (Coords.) (2020). *Avaliación da competencia bilingüe nos idiomas galego e castelán do alumnado de 4º da ESO*. Real Academia Galega. <https://doi.org/10.32766/rag.372>
- LOREDO GUTIÉRREZ, X. e SILVA VALDIVIA, B. (2014). Variables asociadas á competencia gramatical e léxica en galego e castelán do alumnado de Galicia ao final do ensino obrigatorio. *Revista de Investigación en Educación*, 12, 191-208. <https://reined.webs.uvigo.es/index.php/reined/article/view/257>
- MONTOLÍO, E. e SANTIAGO, M. (2000). Objetividad e implicación en el texto académico. En E. Montolío (Coord.), *Manual práctico de escritura académica* (pp. 153-182, vol. 3). Ariel.
- MUÑOZ, N. I. (1999). La desagenticación en el lenguaje científico y en el de la divulgación. *Anclajes. Revista de análisis semiótico del discurso*, 3(3), 95-105. <https://cerac.unlpam.edu.ar/index.php/anclajes/article/view/3208>
- NIETO PAZÓ, A. e TROITIÑO FRAIZ, X. (2015). *Lingua galega e literatura 1º Bacharelato (Aprender é crecer en conexión)*. Anaya.
- PÉREZ GAZTELU, E. e ZULAIKA IJURKO, E. (2014). De dónde partimos. Las prácticas letradas académicas de los estudiantes de bachillerato. En I. Ballano e I. Muñoz (Eds.), *Escribir en el contexto académico* (pp. 53-88). Universidad de Deusto.

- POUDAT, C. e LOISEAU, S. (2005). Autorial presence in academic genres. En E. Tognini-Bonelli e G. Del Lungo-Camiciotti (Eds.), *Strategies in Academic Discourse* (pp. 51-68). John Benjamins. <https://doi.org/10.1075/scl.19.05pou>
- RUBIO, C. F. (2014). Variação na concordância verbal com ‘a gente’ no português europeu. *Fórum Linguístico*, 11(2), 164-177. <https://doi.org/10.5007/1984-8412.2014v11n2p164>
- SALAS VALDEBENITO, M. (2015). Una propuesta de taxonomía de marcadores meta-discursivos para el discurso académico-científico escrito en español. *Revista Signos. Estudios de Lingüística*, 48(87), 95-120. <https://doi.org/10.4067/S0718-09342015000100005>
- SCHERRE, M. M. P., YACOVENCO, L. C. e NARO, A. J. (2018). Nós e a gente no português brasileiro: concordâncias e discordâncias. *Estudos de Lingüística Galega*, 12, 13-27. <https://doi.org/10.15304/elg.ve1.3585>
- SERRANO, M. J. (2013). El pronombre tú como recurso objetivador en español. Variación textual y discursiva. *Borealis. An International Journal of Hispanic Linguistics*, 2(1), 179-197. <https://doi.org/10.7557/1.2.1.2519>
- SERRANO, M. J. (2021a). Gradual objectivity. Variation in the use of the objectivizing second-person singular tú in Spanish. *Journal of Pragmatics*, 176, 44-72. <https://doi.org/10.1016/j.pragma.2021.01.029>
- SERRANO, M. J. (2021b). La función de los constructores de espacio mental en el discurso publicitario: el uso de la segunda persona objetivadora tú. *Oralia*, 24(2), 385-411. <https://doi.org/10.25115/oralia.v24i2.7141>
- SILVA, B., MORÁN, C., BUELA, P. e LOREDO, X. (2008). O ensino da lingua e a literatura galega na ESO. Unha aproximación sociopedagóxica á situación desta área curricular. *Revista galega de educación*, 41, 88-93.
- SILVA VALDIVIA, B. (2004). *Interferencias morfosintácticas do castelán no galego: avaliación da competencia do alumnado de educación secundaria* [Tese de doutoramento inédita]. Universidade de Santiago de Compostela.
- SILVA VALDIVIA, B. (2006). *Castelanismos morfosintácticos no galego. Avaliación da competencia do alumnado*. Universidade de Santiago de Compostela.
- SILVA VALDIVIA, B. (Dir.) (2010). *Avaliación da competencia do alumnado de 4º da ESO nos idiomas galego e castelán*. Universidade de Santiago de Compostela.
- VENEGAS, R., MEZA GUZMÁN, C. e MARTÍNEZ HINCAPIÉ, J. (2003). Procedimientos discursivos en la atribución del conocimiento en tesis de lingüística y filosofía en dos niveles académicos. *RLA. Revista de Lingüística Teórica y Aplicada*, 51(1), 153-179. <http://doi.org/10.4067/S0718-48832013000100008>

A expresión da incerteza nos textos de *CORTEGAL*: un achegamento paradigmático

VÍTOR MÍGUEZ

Universidad del País Vasco / Euskal Herriko Unibertsitatea

1. INTRODUCCIÓN

A capacidade para argumentar é fundamental no desenvolvemento da vida académica e persoal, o cal fai da escritura de textos argumentativos unha tarefa común nos niveis máis elevados da educación (Andrews, 2010, pp. 1-19). A expresión de xuízos epistémicos, isto é, avaliacións de certeza e incerteza, está presente en todo tipo de xéneros discursivos, pero resulta especialmente relevante en textos argumentativos, nos cales é necesario comprender, manexar e xerar argumentos a favor e en contra de puntos de vista en conflito para alcanzar un posicionamento propio (Luna et al., 2020). Existe un grande interese pola expresión da modalidade epistémica en textos argumentativos de aprendices de segundas linguas, onde os textos de falantes non nativos son a miúdo comparados cos de falantes nativos para medir o grao de desenvolvemento das capacidades dos aprendices, así como para propoñer medidas para facilitar a adquisición dos dispositivos lingüísticos que expresan modalidade epistémica (Hu e Li, 2015; McEnery e Kifle, 2002; Yang e Zeng, 2021).

Este traballo ten o duplo obxectivo de contribuír ao coñecemento da expresión da incerteza en textos argumentativos e de afondar no coñecemento do sistema modal epistémico da lingua galega. Con este fin, analizáronse, nos textos argumentativos que constitúen *CORTEGAL*, presentados en Álvarez de la Granja e Rodríguez Rodríguez neste libro, as formas lingüísticas que codifican incerteza seguindo o achegamento paradigmático proposto por Nuyts (2001), aplicado polo autor ás linguas xermánicas occidentais e aínda

non reproducido noutras linguas. Este achegamento parte da idea de que os variados recursos formais con que conta a lingua para expresar a modalidade epistémica están organizados consonte a factores funcionais. En galego, a incerteza pode expresarse mediante, polo menos, verbos auxiliares (por exemplo, *poder*), predicados adxectivos (por exemplo, *é posible*), verbos léxicos (por exemplo, *crer*) e adverbios (por exemplo, *quizais*) (*vid.* Míguez, 2021a). Para dar conta desta variación, consideráronse os factores propostos por Nuyts (2001): a (inter)subxectividade, isto é, o carácter individual ou compartido da responsabilidade sobre o xuízo epistémico; a dispoñibilidade de usos descritivos ou (in)capacidade dos recursos lingüísticos para dar conta dos xuízos expresados por individuos distintos do enunciador; o carácter focal ou non focal da expresión da incerteza; e a estratexia discursiva ou posibilidade de empregar a expresión epistémica para obter certos efectos especiais, como alterar a forza ilocutiva do enunciado ou xestionar a interacción. Xunto a estes factores, tivéronse en conta trazos sintácticos e da estrutura informativa que puidesen influír no uso das expresións epistémicas, como a súa realización en estruturas de complementación ou parentéticas, o seu carácter (non-)temático, a combinación coa negación ou a capacidade de tomar modificadores da expresión epistémica. Preténdese así explicar a variación formal no ámbito da modalidade epistémica en termos de relacións paradigmáticas e contribuír a revelar os factores que guían a escolla entre variantes alternativas para a expresión da incerteza en textos argumentativos escritos no ámbito académico.

O traballo ten a seguinte estrutura: a Sección 2 ilustra os aspectos fundamentais do modelo teórico en que se fundamenta o estudo; a Sección 3 explica o método do estudo, facendo referencia á anotación dos datos e ao tipo de análise estatística empregado; a Sección 4 dá conta dos resultados, tanto de maneira univariada como multivariada; a Sección 5 recolle as principais conclusións e pecha o traballo.

2. O ACHEGAMENTO PARADIGMÁTICO Á MODALIDADE EPISTÉMICA

Na proposta de Nuyts (2001), a modalidade epistémica equivale aos xuízos de probabilidade feitos polo falante en relación cun estado de cousas¹. Trátase dunha categoría conceptual básica que posúe características comúns a outros tipos de cualificacións de estados de cousas, como a modalidade deóntica ou a evidencialidade, pero o seu procesamento cognitivo e a súa

¹ *Estado de cousas* fai referencia ao contido designado por unha predicación, isto é, á estrutura semántica correspondente a algún tipo de suceso do mundo.

expresión lingüística son particulares e altamente complexos, o cal a fai merecente de ser estudada por dereito propio. Neste sentido, Nuyts (2001) interésase polos factores funcionais que motivan a variedade de recursos formais existentes para a expresión da modalidade epistémica e, deste xeito, identifica e compara catro tipos de expresións en alemán, inglés e neerlandés (adverbios e adxectivos modais, verbos cognitivos e auxiliares modais). A distribución destes recursos formais explícase con base en catro factores funcionais:

1. A (inter)subxectividade refírese ao status da evidencia considerada e ao xuízo elaborado a partir dela: un xuízo marcado como subxectivo implica que este é responsabilidade exclusiva do falante, mentres que un xuízo marcado como intersubxectivo supón que a responsabilidade do mesmo é compartida por un grupo amplo de persoas (véxase o capítulo de Silva-Domínguez neste volume para unha visión xeral sobre os recursos de marcado da implicación do falante).
2. A performatividade² ten que ver co feito de que algúns recursos expresivos só serven para manifestar a avaliación modal que fai o falante do estado de cousas no momento da enunciación (uso performativo), mentres que outros tamén poden ser empregados para describir unha avaliación epistémica que non corresponde ao falante no momento da enunciación (uso descritivo).
3. A estrutura informativa, entendida como diferenzas no status ou importancia das partes informativas dun enunciado no contexto discursivo, enfronta a cualificación epistémica ao estado de cousas alcanzado por ela. Segundo este factor, os tipos de expresións modais epistémicas poden ser focais ou non focais.
4. A estratexia discursiva recolle a observación de que as unidades epistémicas son empregadas con frecuencia para obter certos efectos especiais na interacción, como atenuar a aserción ou xestionar a discusión.

Na proposta de Nuyts (2001), estes factores serven para explicar por que os falantes escollen un determinado tipo de expresión fronte aos demais nun contexto determinado. A maneira en que os factores interactúan é complexa, como recolle a Táboa 1.

² Téñase en conta que este uso de *performativo* é netamente diferente do que se fai na influente teoría dos actos de fala (Austin, 1962).

TÁBOA 1. Factores decisivos no uso das expresións epistémicas (Nuyts, 2001, p. 227)

	Cualificación epistémica			
	↓	↓	↓	↓
(inter)subxectividade	-	+	++	(-)
uso descritivo	-	(+)	+	(-)
uso focal	-	++	(+)	(-)
atenuación	-	-	+	-
xestión da discusión	+	-	-	+
	↓	↓	↓	↓
	adverbio	adxectivo	verbo cognitivo	auxiliar

O signo «+» indica que o factor é frecuentemente responsable do uso dun determinado tipo de expresión, «++» indica que o factor é o máis determinante no uso do tipo expresivo, «-» indica que o factor non xoga un papel no uso dese recurso formal e as parénteses arredor dos signos atenúan o seu valor.

Algúns medios expresivos están claramente determinados: os adxectivos son empregados principalmente para a expresión focal da modalidade epistémica, (1), e, en menor medida, para a expresión de xuízos intersubxectivos, (1) e (4a); mentres que os verbos cognitivos se usan principalmente para a expresión subxectiva da modalidade epistémica, (2), e para mitigar elementos do discurso, (3). Ademais, ambos os recursos poden ser empregados para describir cualificacións epistémicas, (4).

- (1) Non só non é seguro, senón que é moi pouco probable que contraten un substituto
- (2) Eu creo que ela nunca confiou en ninguén como en seu avó
- (3) [Son as 11:55 e a reunión comeza ás 12:00. Neste contexto, A dille a B:]
Creo que debemos ir indo
- (4) a. Era moi probable que estivese sendo vixiado
b. O veciño cre que mañá vai chover

Outros tipos de medios expresivos parecen ser neutros en termos da maioría de factores indicados: trátase dos adverbios e dos verbos auxiliares, os cales, segundo Nuyts (2001), non marcan (inter)subxectividade, nin adoitan ser usados para describir cualificacións, nin como foco. O autor si lles outorga un papel na xestión da interacción, lixeiramente diferente á función atenuadora propia dos verbos de cognición. Parece, entón, que estes dous tipos de expresións ocupan un mesmo lugar funcional. Unha das diferenzas que observa Nuyts (2001) é a precisión do valor semántico dos adverbios, moi espe-

cífico en termos da escala epistémica, fronte á maior vaguidade dos auxiliares. Aínda que esta diferenza semántica pode ser relevante na elección dun ou outro elemento, para o autor, a existencia destes dous tipos funcionalmente equivalentes xustifícase principalmente por un feito de natureza sintáctica: a posibilidade de empregar ambos os recursos no mesmo enunciado, de maneira que o falante pode expresar varias dimensións semánticas distribuindo esta carga de maneira proporcional entre diferentes tipos de recursos expresivos.

Todo o anterior leva a Nuyts (2001) a concluír que os diferentes tipos de expresións epistémicas forman un paradigma, é dicir, un conxunto de alternativas formais para a expresión dunha mesma categoría funcional cuxa escolla está determinada por factores alleos a esta última. Así, o autor considera que a modalidade epistémica é inherente á semántica léxica dos distintos tipos de expresións e que os factores funcionais da Táboa 1 actúan (case) sempre a través da sintaxe destes dispositivos formais. O autor explica as diferenzas na especificidade do valor epistémico das clases formais (alta en adxectivos e adverbios, baixa en verbos cognitivos e auxiliares) aludindo a factores externos á categoría funcional da modalidade epistémica: os adxectivos poden ser máis específicos que os adverbios (*cfr. moi/bastante/pouco/moi pouco probable vs. moi/*bastante/*pouco/*moi pouco probablemente*) e estes, á súa vez, sono máis que os verbos cognitivos e os auxiliares. Isto ten que ver coa estrutura informativa, pois unha especificidade alta da cualificación epistémica suxire un status informativo prominente. É dicir, os adxectivos empréganse frecuentemente cando o xuízo epistémico é o foco do enunciado, porque son capaces de soportar a maior carga informativa que esta función require.

Por último, Nuyts (2001, p. 262) observa que as expresións epistémicas se agrupan arredor de dous padróns sintácticos, ilustrados na Táboa 2. Se se consideran, no plano sintáctico, adxectivos e adverbios como variantes dunha mesma clase de expresións, cada unha destas posúe dúas alternativas para a súa realización formal. Na estrutura de complementación, a cualificación epistémica aparece nunha oración superordinada e o estado de cousas nunha oración subordinada. Na estrutura plana, a cualificación «infiltrase» na expresión do estado de cousas (EdC), con diferentes graos de integración, desde moi alta nos auxiliares ata moi baixa nos parentéticos.

TÁBOA 2. Padróns sintácticos das expresións epistémicas (Nuyts, 2001, p. 262)

	Complementación	Estrutura plana
adxectivos/adverbios	<i>é probable que</i> EdC	E <i>probablemente</i> dC
verbos cognitivos	<i>(eu) creo que</i> EdC	E, <i>creo (eu)</i> , dC
auxiliares	<i>pode que</i> EdC	E <i>pode</i> dC

Cada clase expresiva mostra tendencias a respecto destas opcións sintácticas, de acordo cos resultados do estudo: as dúas opcións son comúns con adxectivos e adverbios, os verbos cognitivos teñen preferencia pola expresión parentética e os auxiliares tenden á complementación.

Estas son as liñas mestras do achegamento paradigmático de Nuyts (2001). A meu ver, constitúe unha das explicacións máis sofisticadas da variación na expresión lingüística da modalidade epistémica. Non quere isto dicir que o achegamento paradigmático tal e como o formula Nuyts (2001) sexa un produto impecable. O estudo non foi reproducido noutras linguas, mais é obvio que certas conclusións están determinadas pola mostra de linguas empregada no estudo. Por exemplo, a preferencia dos tipos expresivos por un determinado padrón é algo que moi probablemente variará dunha lingua ou familia de linguas a outra. Neste sentido, Míguez Rego (2021) mostra que os adverbios epistémicos do galego se posicionan na maioría de casos á esquerda do estado de cousas, actuando como temas modais. Este tipo de observacións levan a cuestionar a aplicabilidade do modelo a linguas diferentes daquelas para as que foi concibido, isto é, as xermánicas occidentais. O restante deste traballo está dedicado a comprobar ata que punto o comportamento das expresións epistémicas do galego se axusta ás predicións que ofrece o achegamento paradigmático e a explorar como este pode contribuír á descrición dunha variedade lingüística particular, léase, a linguaxe académica.

3. MÉTODO

Para a elaboración deste estudo, extraéronse de *CORTEGAL* 568 *tokens* de expresións de dúbida, correspondentes aos catro tipos expresivos ilustrados en §2. Dado que *CORTEGAL* é un corpus lematizado, empregouse a busca por lemas para a extracción dos *tokens*. Na selección das observacións, descartáronse manualmente aqueles casos en que non estaba presente a semántica epistémica. Neste sentido, adverbios como *quizais* e *seguramente* ou adxectivos como *probable* tenden fortemente a expresar modalidade epistémica

de maneira exclusiva, mentres que auxiliares como *poder* ou verbos como *crer* ou *pensar* posúen unha gama bastante ampla de significados tanto modais como non modais. No caso de *poder*, das 1684 ocorrencias rexistradas do lema como verbo, só 174, é dicir, aproximadamente o 10%, corresponden á semántica epistémica. Nos casos de *crer* e *pensar*, a porcentaxe de usos epistémicos sitúase por debaixo dun terzo (79/257 ou 30,7%, no caso de *crer*; 99/406 ou 24,38%, no caso de *pensar*). Tamén é relevante sinalar que o adverbio *seica* foi incluído na mostra de datos a pesar de codificar tipicamente valores evidenciais (González-Vázquez, 2021), pois na súa única ocorrencia en *CORTEGAL* é empregado como marcador de posibilidade epistémica, é dicir, como equivalente a *quizais* ou *se cadra*.

Unha vez extraídos e seleccionados os datos, levouse a cabo a anotación manual dos mesmos. Para isto, tivéronse en conta nove variables: catro delas corresponden coas propostas por Nuyts (2001), tratadas en §2, mentres que as cinco restantes están relacionadas con factores da estrutura da información e a súa inclusión pretende compensar as dificultades de facer operativa a distinción entre a expresión focal e non focal da cualificación epistémica.

En galego, a posición non marcada do foco dunha oración é a final (Álvarez e Xove, 2002, p. 78). Ademais, son comúns outros tipos de foco como o contrastivo e o ponderativo. O primeiro destes foi ilustrado no exemplo (1), mentres que o segundo recibe como marca unha maior intensidade tonal na lingua oral e rara vez é marcado na lingua escrita. Así pois, na anotación dos datos decidiuse considerar como casos de expresión focal da cualificación epistémica aqueles en que o marcador modal ocorre ao final da oración (foco non marcado) e aqueles en que se establece un contraste entre marcadores con diferentes valores epistémicos (foco contrastivo). Ningún destes casos se deu nos datos analizados, de maneira que se botou man de factores adicionais, detallados a seguir, para investigar o rol da estrutura informativa na escolla entre expresións epistémicas.

A posición inicial do enunciado adoita considerarse moi sensible desde o punto de vista da estrutura informativa. No caso das expresións de modalidade, a súa ocorrencia na posición inicial supón a súa consideración como temas modais, isto é, como marcas lingüísticas que poñen de manifesto a actitude do emisor ante a súa mensaxe e serven como punto de partida para a organización da mesma (Halliday, 2014, pp. 105-112; Hoye, 1997, pp. 148-152). Na anotación dos datos, cando a expresión epistémica ocupa a posición inicial do enunciado é considerada tema.

Canto aos padróns sintácticos ilustrados na Táboa 2, consideráronse as estruturas de complementación (*é probable que, creo que, pode que*) fronte á non complementación.

Outro factor, fundamental desde o punto de vista informativo, é a polaridade. A expresión lingüística da negación é de seu contrastiva, pois un enunciado negativo (case) sempre constitúe unha reacción ao seu equivalente afirmativo, estea este explícito ou non (Nuyts, 2001, p. 80). Dado que o entrar en contraste é unha razón de peso para outorgar prominencia informativa a un elemento lingüístico (Chafe, 1994, pp. 76-78), os datos foron anotados tendo en conta a súa polaridade (afirmativa ou negativa).

As expresións epistémicas poden afectar ao enunciado no seu conxunto ou só a unha parte del. Neste último caso, a unidade afectada pode conter información prominente, situada no primeiro plano do discurso, ou información secundaria, relevante para os fins comunicativos do emisor pero situada nun lugar pouco destacado. Seguindo a visión tradicional (e.g., Givón, 2001, vol. 2), considerouse que o marcador epistémico afecta a información secundaria cando aparece en oracións subordinadas adverbiais e relativas explicativas ou en aclaracións parentéticas.

A posibilidade de modificar a expresión epistémica xa foi mencionada anteriormente e distingue os adxectivos e, en certa medida, os adverbios, con bastante flexibilidade á hora de proporcionar especificacións adicionais da cualificación epistémica, dos auxiliares e os verbos cognitivos, cunhas posibilidades moito máis limitadas neste sentido.

Os datos anotados foron sometidos a unha análise estatística exploratoria mediante o método da Análise de Correspondencia Múltiple (ACM) (*vid.* Greenacre, 2007; Levshina, 2015, pp. 367-385). Este método permite converter os padróns de variación en distancias ao longo de varias dimensións e así capturar asociacións entre variables. Desta maneira, os niveis das variables son representados como puntos nun mapa de dúas dimensións, o cal axuda a visualizar e interpretar os resultados cando está presente un número elevado de variables explicativas. Para levar a cabo a ACM, empregouse a versión máis recente do contorno R (R Core Team, 2022) e o seu paquete FactoMineR (Lê et al., 2008).

4. RESULTADOS

Esta sección dá conta dos resultados atendendo, por unha parte, a cada unha das variables, isto é, (inter)subxectividade, performatividade, estrutura informativa e estratexia discursiva (§4.1), e, por outra parte, á súa contribución

conxunta á variación observada mediante unha análise multifactorial baseada no método ACM (§4.2).

4.1. Análise univariada

O número de usos epistémicos de adverbios, auxiliares e verbos cognitivos en *CORTEGAL* é moi parello (Táboa 3): 188 ocorrencias no caso dos adverbios; 179 no caso dos verbos de predicado mental; e 174 no caso dos auxiliares. Os adxectivos mostran unha incidencia moito menor, con 35 ocorrencias de usos epistémicos.

TÁBOA 3. Distribución de *tokens* e lemas por tipo de expresión

	Adverbios	Adxectivos	Auxiliares	Verbos	TOTAL
# <i>tokens</i>	188	35	174	179	576
# lemas	11	3	1	3	18

Como mostra a Táboa 3, as diferenzas no número de lemas por categoría son notables. Atopáronse once lemas diferentes para os adverbios (*ao mellor, igual, posiblemente, probablemente, quizá, quizais, quizás, se cadra, seguramente, seica e talvez*), mentres que o resto de categorías non superan os tres lemas: *coidar, crer e pensar*, no caso dos verbos cognitivos; *posíbel, posible e probable*, no caso dos adxectivos; e *poder* no caso dos auxiliares. Os lemas, exemplos dos mesmos e as súas respectivas frecuencias poden consultarse na Táboa 4 do apéndice. Os resultados do estudo de corpus por variable están recollidos na Táboa 5 do apéndice. Na exposición que segue, ofrécese un resumo dos resultados acompañado de exemplos ilustrativos de cada categoría de análise.

4.1.1. (Inter)subxectividade

Os catro tipos de expresións examinados difiren notablemente no relativo á (inter)subxectividade. Os verbos cognitivos son sempre subxectivos, isto é, producen unha cualificación epistémica que ten como único responsable o emisor³.

- (5) A gastronomía e a cociña acadaron moita popularidade. Eu *creo* que isto débese ó turismo

³ Os exemplos introducidos a partir de aquí están tirados de *CORTEGAL* e corresponden, non sendo que se indique o contrario, á versión final do estudante. As expresións epistémicas están destacadas en itálica.

De feito, a semántica subxectiva está tan ligada a este tipo expresivo que, con frecuencia, é este o único significado expresado polo verbo (e, nese caso, a observación correspondente foi excluída da mostra). Neste sentido, é revelador observar aqueles exemplos en que o predicado de estado mental coincide cunha locución de punto de vista subxectivo, como *na miña opinión*, pois a estandarización discursiva que ofrece *CORTEGAL* (véxase §2.5 en Álvarez de la Granja e Rodríguez Rodríguez nesta obra) elimina o verbo cognitivo nestes casos. Isto é coherente coa observación de que o uso do verbo é puramente subxectivo, sen pegada da semántica epistémica.

- (6) a. Ao meu parecer non *creo* adecuado meter dentro do mesmo saco a profesión de futbolista coa de modelo (*versión estudiante*)
 b. Ao meu parecer, non é adecuado meter dentro do mesmo saco a profesión de futbolista e a de modelo (*estándar discursivo*)

Os adxectivos mostran o padrón oposto ao dos verbos: excepto dous, todos os seus usos son intersubxectivos, é dicir, presentan a cualificación epistémica como compartida entre o emisor e outros individuos.

- (7) En conclusión, a popularidade da gastronomía responde a unha serie de fenómenos complexos ligados ó cambio da sociedade, e *é posible* que vaia manterse durante moito tempo

Os únicos casos de usos subxectivos dun adxectivo débense á presenza explícita da primeira persoa no contexto inmediato. Nestes exemplos, trátase da locución de punto de vista *na miña opinión* e do verbo *ver* en primeira persoa.

- (8) a. Na miña opinión, *é moi posible* que este tipo de comida pase de moda, pero o que seguro non vai pasar de moda é a gran popularidade da cociña
 b. Tamén dase o caso de non sexa simplemente unha moda e definitivamente estableza un cambio cultural pero na miña opinión o vexo inviable xa que na sociedade na que vivimos infestada pola comida lixo *vexo pouco probable* que o conxunto da sociedade cambie xa que a opción de que sexa unha simple moda é máis razoable e comprensible

Fronte á clara especialización de verbos e adxectivos nunha das dimensións da (inter)subxectividade, adverbios e auxiliares mostran ambivalencia nesta coordenada semántica. Os dous tipos expresivos teñen unha proporción considerable de usos subxectivos e intersubxectivos, aínda que os adverbios mostran unha preferencia pola subxectividade (69,68%) e os auxiliares pola intersubxectividade (55,17%). As series (9) e (10) ofrecen exemplos de cada unha das categorías.

- (9) a. A todo isto, non pretendo afirmar que non hai traballo de fondo, de feito, hai moitas horas de esforzo e traballo, pero iso non o mostran; os rigurosos e duros adestramentos diarios, todas as veces que se cometen erros e hai que repetir o exercicio ou dar outra volta para afianzar terreo. Todos esos momentos non se mostran e *igual* sería moi interesante facelo, porque nesta vida, de balde, non hai nada (*adverbio, subxectivo*)
- b. A necesidade ou entusiasmo de converter o noso mundo nun país máis satisfeito trouxo consigo principalmente novidades no mundo da cociña; respecto a elaboración de novos pratos... Esto produciu un cambio cultural que *posiblemente* se verá incrementado co paso dos anos e potenciáronse os coñecementos adquiridos polos diferentes territorios do mundo (*adverbio, intersubxectivo*)
- (10) a. Na miña opinión, unha das causas de esta popularidade repentina *pode* ser a cantidade de programas televisivos de cociña que dan na televisión últimamente (*auxiliar, subxectivo*)
- b. Polo tanto, o mundo gastronómico aumenta por partida dobre, por un lado, grazas ao desenvolvemento do turismo e polo outro, debido á globalización e a aparición de novos programas de entretemento, polo que este último *podería* ser un crecemento paxaxeiro, que se esgote a medida que perdan popularidade este tipo de programas (*auxiliar, intersubxectivo*)

A (inter)subxectividade dos usos dunha expresión é determinada pola súa sintaxe (Nuyts, 2001). Así, os verbos cognitivos adoitan aparecer en construcións persoais e expresan, en consecuencia, modalidade epistémica subxectiva; os adxectivos ocorren en construcións impersoais e os seus usos son, polo xeral, intersubxectivos; e adverbios e auxiliares participan de estruturas non marcadas e a (inter)subxectividade dos seus usos mostra unha dependencia maior doutros factores contextuais.

Neste sentido, Delbecque (2009) sinala que en español existe un contraste entre os adxectivos *cierto* e *seguro* en termos de (inter)subxectividade, o cal está relacionado cos padróns sintácticos (*es cierto* vs. *estoy seguro*). Con todo, a mesma autora mostra que a oposición entre os adxectivos non depende unicamente dos padróns sintácticos, pois tamén se dá cando aqueles aparecen na mesma construción. Isto suxire que a (inter)subxectividade pode ser parte do significado léxico dunha expresión. Un argumento a favor desta idea é o feito de que os adverbios derivados destes adxectivos mostran o mesmo contraste, aínda en diferentes linguas románicas. Para o italiano, Pietrandrea (2008, p. 242) indica que *certamente* funciona como un «polyphonic trigger», mentres que *sicuramente* expresa «monophonic judgments». De maneira similar, Míguez (2021b, p. 143) atribúe o diferente desenvolvemento diacrónico de *certamente* e *seguramente* en galego ao contraste en termos de (inter)subxectividade entre os dous adverbios.

4.1.2. *Performatividade*

O uso por defecto dunha expresión epistémica é performativo. Como no caso da (inter)subxectividade, a manifestación desta dimensión semántica depende da sintaxe mediante a que é expresada. Así, adxectivos e verbos cognitivos, ao constituír predicados con soporte verbal, permiten usos descritivos estruturais que atribúen a cualificación epistémica a entidades diferentes do emisor, fronte a adverbios e auxiliares, que dependen doutras construcións, como as reportativas, para alcanzar lecturas descritivas. Destas observacións deriva a idea de que a performatividade xoga un papel no uso de adxectivos e verbos (Táboa 2). Porén, no presente estudo, só estes últimos presentan usos descritivos, cunha incidencia inferior ao 7% (12/179). Para o resto de expresións non foron detectados usos deste tipo.

- (11) Moita xente *pensa* que é unha moda pasaxeira e dentro de pouco desaparecerá, pero é verdade? Na miña opinión, non

4.1.3. *Estrutura informativa*

O medio escrito dificulta a detección de usos focais das expresións epistémicas, pois, como se sinalou máis arriba (§3), o principal xeito de marcaxe do foco é a intensidade tonal, que xeralmente carece de reflexo na escrita. Non obstante, non se detectaron casos de foco non marcado ou de foco contrastivo, o cal fai supoñer que, polo menos na lingua escrita, os usos focais das expresións epistémicas son moi infrecuentes. Isto non é sorprendente, xa que a cualificación epistémica, pola súa propia natureza, tende a ser información secundaria en relación co estado de cousas que modifica.

Outras variables da estrutura informativa resultaron máis fáciles de operativizar. É o caso do tema, onde se observan diferenzas notables entre tipos expresivos. Os verbos cognitivos teñen a maior preferencia pola función temática (56,42%), seguidos dos adxectivos (37,14%), dos adverbios (26,06%) e dos auxiliares (13,79%).

- (12) a. *Penso* que a globalización ten moito que ver coa extensión do fanatismo, no bon sentido da palabra, pola gastronomía
 b. *É probable* que na sociedade actual o que esta de moda agora deixe de estalo nuns anos, xa que vivimos nunha sociedade de cambios moi rápidos
 c. *Quizáis* esta familia non vexa necesario gastar máis en alimentación porque considera que a calidade do produto non varía tanto
 d. *Pode* que esta crecente preocupación polo mundo culinario sexa soamente moda

Os exemplos a, b e d da serie (12) exemplifican a complementación para os verbos, os adxectivos e os auxiliares, respectivamente. Aínda que os usos complementadores dos adverbios son posibles (por exemplo, *seguramente que*), non hai ningún caso na mostra de datos. A complementación parece estar relacionada coa función temática, pois os tipos expresivos que posúen maior número de usos temáticos, isto é, verbos e adxectivos, mostran unha clara preferencia por padróns de complementación (> 98%). Con todo, os adverbios, non tendo usos complementadores, posúen unha proporción maior de usos temáticos que os auxiliares, os cales si mostran recorrencia na complementación (15,52%).

Canto aos usos sen complementación, estes son a única opción para os adverbios, son dominantes nos auxiliares (84,48%), moi minoritarios nos verbos (1,12%) e inexistentes nos adxectivos. As series (9) e (10) ilustran a non complementación para adverbios e auxiliares, respectivamente. Os dous usos sen complementación detectados para os verbos son parentéticos e recóllense en (13).

- (13) a. Resultan moi interesantes a experimentación e a arte gastronómicas, pero isto é unha empresa que, *creo*, se debería realizar cando todos teñan xa satisfeita esa necesidade vital que é comer
 b. Grazas a isto permítesenos coñecer polo gusto e o paladar, certas culturas que ata hace uns anos descoñecíamos. Pero isto, *penso*, tamén ten a súa parte negativa

Os usos parentéticos marcados entre comas son mostra dun claro deslindamento prosódico entre o marcador epistémico e o estado de cousas que modifica. No caso dos adverbios, hai poucos exemplos claros de usos parentéticos, isto é, marcados con comas (14a), aínda que moitos permitirían esta lectura (14b). Porén, hai algúns casos en que a lectura parentética é dificilmente aceptable (14c).

- (14) a. Pode tratarse, *quizais*, dunha moda pasaxeira
 b. Sen embargo, esta moda de modernizar a comida é *probablemente* algo pasaxeiro
 c. O que *seguramente* pase con este tema tan falado nestes últimos anos, sexa o que pasa con todas as modas, que se deixen a un lado e que os restaurantes volvan a servir para o mesmo que antes, para comer

Para os auxiliares non se detectou ningún posible uso parentético, o cal pode tomarse como indicativo de seren o tipo expresivo co maior grao de integración no estado de cousas.

No terreo da polaridade tamén se dan contrastes salientables entre os tipos de marcadores. A proporción de usos negativos é próxima no caso de adverbios, adxectivos e verbos (entre o 17,14% e o 23,46%), mentres que para os auxiliares só se detectou un caso (0,57%), recollido en (15).

- (15) Pero, é sostenible esta gastronomía na que todo o mundo pretende coñecer?
Pode que si, ou *pode* que non

A relación entre os auxiliares modais e a negación é complexa. A combinación do auxiliar *poder* e os seus equivalentes noutras linguas con partículas negativas favorece lecturas non epistémicas do auxiliar, nomeadamente, lecturas de imposibilidade dinámica (por exemplo, incapacidade física) ou deóntica (por exemplo, inaceptabilidade moral) (cfr. Nuyts, 2001, pp. 214-218; Palmer, 1990, pp. 60-62). A excepción a esta forte tendencia é o padrón de complementación, onde é máis común atopar casos de negación na cláusula subordinada (Nuyts, 2001, p. 215). Os resultados do presente estudo son coherentes con estas observacións, pois, como se pode observar en (15), o único caso de negación encontrado corresponde ao padrón de complementación.

Os catro tipos de expresións aparecen en contextos pouco prominentes e, neses casos, afectan información secundaria no enunciado. Os adverbios son os que con maior frecuencia ocorren nestes contextos (23,94%), seguidos dos adxectivos (14,29%), os auxiliares (12,64%) e os verbos (4,47%).

- (16) a. Pola contra, *igual* son moi optimista, sí creo que podemos cambiar as nosas mentalidades
b. Pero se a demanda é moi grande, certas empresas poden aproveitarse e subir os prezos, de modo que o seu consumo xa non estará á disposición de calquera, e polo tanto pasarase a comprar outros produtos de prezos menores, entrando no consumo outras marcas e empresas, que *probablemente* acaben facendo o mesmo
c. Son, xa que logo as dúas caras da mesma moeda moneda, e xa que non é *posible* que esta mentalidade cambie, si que debemos reflexionar e aplicar a tan conveniente expresión: nin tanto, nin tan pouco
d. Na miña opinión, o tempo que dure esta moda dependerá da imaxinación e das innovacións que haxa na gastronomía posto que coma con todo chega un momento no que as cousas que sempre siguen igual cánsante, aínda que eu *penso* que será un cambio cultural duradeiro

Os tipos de contextos con información secundaria son bastante variados: (16a) mostra o uso dun adverbio nun inciso parentético; (16b) mostra un adverbio nunha oración subordinada relativa explicativa; e (16c-d) recollen,

respectivamente, un adxectivo e un verbo en subordinadas adverbiais. Un exemplo disto último cun auxiliar modal témolo en (10b).

O último dos factores relativos á estrutura informativa é a capacidade de tomar modificadores que especifiquen de maneira adicional a cualificación epistémica. Os resultados axústanse ao adiantado en §3. Aínda que é posible modificar os usos epistémicos de auxiliares e verbos (por exemplo, mediante o adverbio *ben*), non hai un só caso na mostra de datos. Canto aos adverbios, estes son modificados en tan só tres ocasións (1,6%) e sempre polo adverbio *moi*, que combina con *probablemente* en dous casos e con *posiblemente* nun. Fronte ao escaso ou nulo grao de modificación do resto de expresións, os adxectivos toman modificadores na maioría das súas aparicións (51,43%). Hai seis casos de *moi probable*, un de *pouco probable*, un de *moi pouco probable*, sete de *máis probable*, dous de *moi posible* e un de *máis posible*⁴. Entre os lemas adxectivos as diferenzas son notables: *probable* é modificado en máis do 70% das súas ocorrencias, fronte a *posible*, que é modificado en menos do 25% dos casos. Esta diferenza entre adxectivos pódese explicar polo feito de que expresan dimensións epistémicas diferentes, unha graduable no caso de *probable* e unha non graduable no caso de *posible* (Rodríguez-Españeira, 2010, pp. 197-198).

4.1.4. *Estratexia discursiva*

Todas as clases de expresións estudadas son susceptibles de dous tipos de usos estratéxicos: a atenuación e a concesión. A atenuación é unha dimensión pragmática que recibiu grande atención nas últimas décadas (e.g., Caffi, 2007) e a súa identificación fica ligada á implicación das imaxes públicas dos participantes (Albelda Marco e Estellés Arguedas, 2021, pp. 77-80). Desta maneira, empregar unha estratexia de atenuación ten como fin protexer unha imaxe pública, sexa esta a do emisor, a do(s) interlocutor(es) ou a dunha terceira persoa. Como consecuencia, o estudo da atenuación é moi frutífero en contextos orais conversacionais, propiamente interactivos, aínda que tamén é perfectamente viable en xéneros escritos de diverso tipo (*vid.* Albelda Marco, 2018; Míguez, 2022).

En *CORTEGAL*, a función atenuadora dos marcadores epistémicos diríxese a salvagardar a imaxe propia, o cal é consecuente co carácter monolóxico propio dos textos argumentativos que conforman o corpus. Mediante o uso de expresións epistémicas en exemplos como (16a), a persoa autora

⁴ Exemplos en contexto correspondentes a cada unha das expresións poden consultarse na Táboa 6 do apéndice.

do texto intenta mitigar a imaxe de inocencia que se podería derivar da súa afirmación; a xustaposición de *creo* e *quizais* en (17a) serve para suavizar a forza da propia opinión e cultivar así unha imaxe de humildade; isto mesmo acontece en casos como (10b) e (17b), en que *poder* epistémico combina co pospretérito con valor temporal presente, un coñecido mecanismo de atenuación (Álvarez e Xove, 2002, p. 281; Freixeiro Mato, 2006, p. 353).

- (17) a. Certo é que este auxe da comida ven acompañado dunha subida nos seus prezos, debido á súa costosa elaboración, polo que non é un fenómeno ao alcance de todo o mundo, senón soamente ao alcance duns poucos. Por esta banda *creo* que *quizais* non estea ben plantexado, polo feito de que debería ser algo ao que todos poidésemos acceder aínda que só fose para probalo
 b. E canto máis se produce, máis se gaña, pero tamén máis se consume, xerando unha necesidade de consumo nun principio inexistente. Isto *podería* ser un dos motivos polos que a sociedade vaise facendo cada vez máis superficial

Desde o punto de vista formal, a concesión é un padrón sintáctico en que unha oración subordinada enuncia un feito que contrasta co enunciado na oración principal sen chegar a invalidalo (Álvarez e Xove, 2002, pp. 163-166). Con todo, Couper-Kuhlen e Thompson (2000) enténdena como unha estratexia pragmática que comprende tres partes secuenciais: un primeiro falante defende unha idea X e un segundo falante reconece (concede) a validez de X pero introduce unha idea Y potencialmente contrastiva. Así, a concesión non se limita ao uso de conxuncións subordinantes concesivas, como *aínda que*, senón que comprende tamén outros padróns formais en que se establece un contraste entre dous argumentos. Neste tipo de contextos, os marcadores epistémicos de dúbida introducen a idea concedida X, suavizando, deste xeito, o desacordo vindeiro Y, que se presenta como un contraste a X mediante unha conxunción adversativa. O feito de que X sexa introducida mediante un marcador de incerteza revela o baixo compromiso do emisor con X e anuncia a idea discrepante Y. Nos textos de *CORTEGAL*, a idea concedida X é normalmente un argumento presentado no texto que aparece na proba ABAU como referencia para desenvolver a argumentación arredor do tema proposto.

- (18) a. *Ao mellor* isto da gastronomía será unha moda pasaxeira, como todas as modas, pero se foi acadando esta popularidade ao longo dos anos, por algo é, e polo tanto esperemos que dure bastante porque é unha axuda para a nosa economía
 b. Na miña opinión, *é moi posible* que este tipo de comida pase de moda, pero o que seguro non vai pasar de moda é a gran popularidade da cociña
 c. En definitiva, esta devoción e gusto pola cociña *pode* que en parte chegase

como unha moda, pero sen dúbida non como as outras, porque esta non será unha moda pasaxeira

d. En cuanto á produción, *creo* que si somos moi produtores pero opino que non é algo que nos guste moito, falando de maneira xeral

A distribución de usos estratéxicos entre os tipos de expresións é irregular. Os auxiliares modais teñen a porcentaxe máis alta (31,03%) e mostran certa preferencia pola concesión (17,24%) fronte á atenuación (13,79%). Os adxectivos, con moitos menos usos estratéxicos (8,57%), mostran a mesma preferencia que os auxiliares (5,71% fronte a 2,86%). Os adverbios e os verbos teñen a menor proporción de usos estratéxicos (4,79% e 2,79%, respectivamente) e unha porcentaxe moi parecida, en ambos os casos, de usos atenuadores e concesivos.

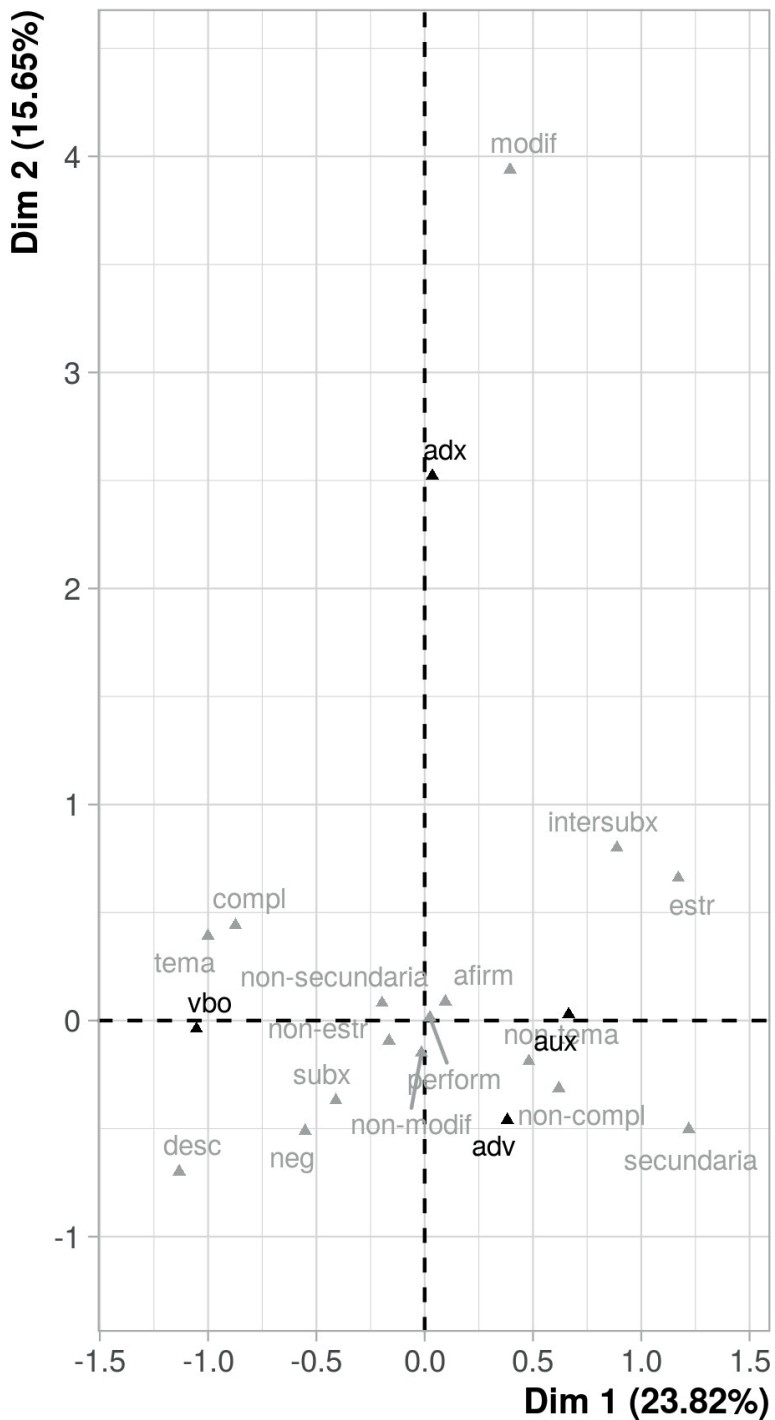
4.2 Análise multifactorial

A Figura 1 contén o mapa da ACM, que servirá de base para interpretar os resultados desde un punto de vista multivariado. O mapa contén dous eixes: o horizontal representa a primeira dimensión e o vertical a segunda. As Táboas 7 e 8 do apéndice mostran o valor R^2 , que mide as contribucións estatisticamente significativas á configuración de cada dimensión.

A primeira dimensión opón as expresións epistémicas principalmente consonte a trazos sintácticos: na metade esquerda do mapa atopamos usos temáticos e con complementador, que claramente corresponden cos verbos cognitivos; na metade dereita sitúanse os usos non temáticos e sen complementador, que se asocian aos adverbios e aos auxiliares. A esta dimensión tamén contribúen en boa medida as oposicións subxectivo / intersubxectivo, información non-secundaria / información secundaria e uso non estratéxico / uso estratéxico, que contrastan, de novo, os verbos cognitivos cos adverbios e os auxiliares. Contribúen de maneira significativa, aínda que cun peso moito menor, os usos negativos e descritivos, asociados aos verbos cognitivos, mentres que os usos afirmativos e performativos se sitúan moi próximos ao punto central do mapa e poden considerarse trazos non marcados.

En termos da primeira dimensión, os adxectivos ocupan un lugar central, representando así un punto medio entre os dous conxuntos de expresións anteriores en relación cos factores relevantes. Porén, o que claramente singulariza os adxectivos con respecto ao resto de expresións é a segunda dimensión, onde a variable con maior peso é a presenza ou ausencia de modificadores. Mentres que os adxectivos están fortemente asociados á presenza de modificadores que especifican de maneira adicional a cualificación epistémica,

FIGURA 1. Mapa da Análise de Correspondência Múltipla (ACM)



o resto de expresións, incluídos os adverbios, está asociado á ausencia de modificadores, trazo situado na zona central da Dimensión 2. A esta dimensión tamén contribúe a (inter)subxectividade, factor que opón os adxectivos, maioritariamente intersubxectivos, ao resto de expresións, para as cales a semántica subxectiva ten un peso variable, pero que, en calquera caso, representa unha parte relevante dos seus usos. O resto de variables contribúen de maneira pouco destacada á configuración desta dimensión.

É interesante observar a situación de cada un dos puntos correspondentes aos catro tipos de expresións epistémicas. Os adxectivos e os verbos cognitivos son os máis afastados do punto central e dos outros tipos de expresións, o cal indica que teñen un perfil funcional moi especializado. Os adxectivos, como xa se adiantou, ofrecen o soporte ideal para levar a cabo especificacións adicionais da cualificación epistémica, pois teñen a capacidade de combinar facilmente con adverbios de cantidade. Esta maior carga informativa depositada no elemento cualificador pode entenderse como unha mostra de maior peso discursivo do xuízo epistémico e parece apoiar a idea de Nuyts (2001) segundo a cal o uso dos adxectivos vén determinado por factores relacionados coa estrutura informativa. Con todo, non se puideron identificar usos focais claros, mentres que outros factores da estrutura informativa, como o alcance sobre información prominente ou secundaria, xogan un papel menor á hora de distinguir entre tipos expresivos.

Os verbos cognitivos opóñense a adverbios e auxiliares fundamentalmente en termos sintácticos: predicados como *coidar*, *crer* e *pensar* mostran unha forte tendencia a actuar como temas modais do enunciado e a aparecer en padróns de complementación. Nisto coinciden en boa medida cos adxectivos. Porén, os verbos cognitivos están especializados nos usos subxectivos e, en menor medida, descritivos, en liña coa proposta de Nuyts (2001).

Por último, os puntos correspondentes a adverbios e auxiliares encóntanse moi próximos no mapa. Hai, polo tanto, varios trazos que caracterizan estes dous tipos de expresións. En primeiro lugar, aínda que funcionan a miúdo como temas modais (*vid.* Táboa 5), son as expresións que con menor frecuencia se sitúan ao comezo do enunciado. En segundo lugar, fronte a adxectivos e verbos cognitivos, adverbios e auxiliares rara vez aparecen en padróns de complementación. En último lugar, os dous tipos de expresión tamén mostran asociacións con trazos pouco típicos, é dicir, afastados do centro do mapa, a saber, o alcance sobre información secundaria e a participación en estratexias discursivas. Aínda que ambos os recursos están asociados a estas dúas variables, estas xogan un certo papel á hora de distinguir

un tipo expresivo do outro, pois as asociacións son moito máis fortes entre os auxiliares e os usos estratéxicos e os adverbios e a información secundaria. A trazos grosos, estes resultados tamén están en liña coas ideas de Nuyts (2001), quen considera adverbios e auxiliares formas equivalentes desde o punto de vista funcional.

Se os presentes resultados apoian, en liñas xerais, a proposta de Nuyts (2001) no referente aos factores determinantes no uso das expresións epistémicas (Táboa 1), os mesmos resultados levan a cuestionar a validez da súa proposta sobre a sintaxe destas expresións (Táboa 2). É certo que adverbios e adxectivos mostran unha clara complementariedade canto ao tipo de estrutura sintáctica en que adoitan aparecer; porén, os verbos cognitivos e os auxiliares non mostran unha ambivalencia neste sentido: os primeiros aparecen en padróns de complementación na inmensa maioría dos casos, mentres que os segundos prefiren de maneira clara a estrutura sen complementación. Aínda máis, estes padróns son opostos aos dos resultados de Nuyts (2001), onde os verbos cognitivos prefiren os usos parentéticos e os auxiliares a complementación.

5. CONCLUSIÓNS

Os resultados deste estudo mostran que os catro tipos de expresións epistémicas estudados (adverbios, adxectivos, auxiliares modais e verbos cognitivos) se axustan a tres perfís funcionais claramente diferenciados, con adverbios e auxiliares representando variantes dun mesmo perfil. Os adxectivos son o tipo expresivo cun perfil máis especializado, ligado fundamentalmente á modificación adicional da cualificación epistémica. Este aspecto explica que os adxectivos sexan as expresións epistémicas menos frecuentes na mostra de datos, pois a súa alta especificidade semántica implica, desde o punto de vista informativo, un lugar central do xuízo epistémico, que pola súa propia natureza adoita constituír información secundaria con respecto ao estado de cousas. Os adxectivos tamén están relacionados coa intersubxectividade e, de novo, isto pode dar conta da súa baixa frecuencia, dado que o xénero argumentativo favorece a expresión subxectiva da dúbida (na mostra de datos temos un 69% de usos subxectivos fronte a un 31% de usos intersubxectivos).

Este carácter subxectivo do xénero textual examinado explica a prevalencia de verbos cognitivos como *crer* e *pensar*, que conforman, en virtude da súa sintaxe número-persoal, o molde ideal para a expresión da dúbida subxectiva. Ao igual que no caso dos adxectivos, o seu uso é sensible a factores da estrutura informativa, pois os verbos cognitivos mostran unha forte

tendencia a actuar como temas do enunciado e a aparecer en estruturas de complementación.

Adverbios e auxiliares son os tipos de expresións cun uso menos dependente da estrutura informativa: mostran unha asociación moito menor coa función temática e a estrutura de complementación que os adxectivos e os verbos cognitivos. Ademais, os adverbios están asociados co alcance sobre información secundaria ou pouco prominente e os auxiliares son adoito empregados con fins estratéxicos. Trátase de dispositivos lingüísticos flexibles, que permiten lecturas tanto subxectivas como intersubxectivas e teñen facilidade tanto para integrarse no enunciado como para desempeñar función temática.

A trazos largos, os resultados do presente estudo no referido ao carácter paradigmático das expresións epistémicas son coherentes coas ideas propostas por Nuyts (2001): a (inter)subxectividade, diversos factores da estrutura informativa e a estratexia discursiva explican a distribución dos recursos formais dispoñibles para a transmisión da semántica epistémica. É notable aquí a ausencia da performatividade, que segundo Nuyts (2001) sería determinante no uso de verbos cognitivos e adxectivos. Só se atoparon usos descriptivos para os verbos e nun número moi reducido, o cal fai difícil defender que este factor xogue un papel central no uso deste tipo de expresións.

Sobre o uso de marcadores epistémicos en textos argumentativos académicos, o presente estudo permite chegar a dúas conclusións principais. En primeiro lugar, prevalecen os usos subxectivos, performativos e temáticos dos marcadores epistémicos. Isto apunta cara á intención das persoas autoras de presentar as hipóteses e opinións propias como puntos de partida para a construción do texto e revela o lugar central que a asignación da responsabilidade sobre a información ocupa na argumentación. En segundo lugar, son claramente dominantes os usos non estratéxicos dos marcadores epistémicos. Polo tanto, a presenza destes dispositivos lingüísticos responde máis á necesidade de formular hipóteses xenuínas sobre a información empregada na argumentación que a conseguir efectos especiais sobre o discurso, como atenuar ou conceder. Estes usos estratéxicos non son raros entre os marcadores epistémicos e responden á necesidade de protexer a aparencia propia cultivando unha imaxe de humildade, pero en ningún caso serven para explicar o amplo uso de marcadores epistémicos en textos argumentativos.

O principal obxectivo dun estudo de gramática descritiva coma este é coñecer a realidade textual dunha determinada coordenada do sistema lingüístico. Con todo, o presente traballo suxire liñas de investigación de natureza

didáctica que sería interesante abordar en futuros traballos. As investigacións futuras sobre a materia beneficiaríanse dun interese pola competencia individual do alumnado. No conxunto dos textos de *CORTEGAL*, atopamos a maioría dos recursos para a expresión da incerteza en galego. Isto suxire que existe, entre o conxunto do estudantado, un dominio das principais maneiras de expresar dúbida, pero conviría comprobar en cada texto a dispersión sintáctica (a combinación de recursos de diferente natureza, coma adxectivos e verbos) e léxica (por exemplo, o uso dunha diversidade de adverbios), co fin de achegarse á competencia individual do alumnado. Canto á dispersión sintáctica, este traballo ofrece unha explicación da menor frecuencia dos adxectivos e o uso parello do resto de recursos no conxunto do corpus (véxase a síntese ao respecto ao comezo desta sección), pero non analiza o uso que cada texto fai dos diferentes tipos expresivos. Canto á dispersión léxica, a Táboa 4 evidencia que existe unha clara preferencia por un número reducido de formas, mesmo cando existe unha gran variedade de alternativas, como no caso dos adverbios. Non sorprende que as formas que resultan máis desfavorecidas sexan aquelas non coincidentes co castelán, como *se cadra* para os adverbios e *coidar* para os verbos, un ben coñecido padrón de selección léxica mediado pola lingua dominante. Desde o punto de vista didáctico, é fundamental fornecer os aprendentes dun conxunto de ferramentas amplo que lles permita construír con éxito textos argumentativos. Neste sentido, a investigación sobre a gramática dos marcadores epistémicos debe colocarse no contexto máis amplo da reflexión sobre os recursos que interveñen no proceso de argumentación e estar ao servizo da superación das dificultades de produción e comprensión dos textos. Ademais, non pode ser allea ás dinámicas do contacto lingüístico.

REFERENCIAS BIBLIOGRÁFICAS

- ALBELDA MARCO, M. (2018). La variación genérico-discursiva de la atenuación como resultado de la variación de la imagen. *Spanish in Context*, 15(2), 346-368. <https://doi.org/10.1075/sic.00018.alb>
- ALBELDA MARCO, M. e ESTELLÉS ARGUEDAS, M. (2021). Mitigation revisited. An operative and integrated definition of the pragmatic concept, its strategic values, and its linguistic expression. *Journal of Pragmatics*, 183, 71-86. <https://doi.org/10.1016/j.pragma.2021.07.002>
- ÁLVAREZ, R. e XOVE, X. (2002). *Gramática da lingua galega*. Galaxia.
- ANDREWS, R. (2010). *Argumentation in higher education: Improving practice through theory and research*. Routledge. <https://doi.org/10.4324/9780203872710>

- AUSTIN, J. L. (1962). *How to do things with words*. Oxford University Press.
- CAFFI, C. (2007). *Mitigation*. Elsevier.
- CHAFE, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. The University of Chicago Press.
- COUPER-KUHLEN, E. e THOMPSON, S. A. (2000). Concessive patterns in conversation. En E. Couper-Kuhlen e B. Kortmann (Eds.), *Cause—Condition—Concession—Contrast: Cognitive and discourse perspectives* (pp. 381-410). Mouton de Gruyter. <https://doi.org/10.1515/9783110219043.4.381>
- DELBECQUE, N. (2009). Acerca de la relación entre *cierto* y *seguro*. En M. Veyrat Rigat e E. Serra Alegre (Eds.), *La lingüística como reto epistemológico y como acción social. Estudios dedicados al Profesor Ángel López García con ocasión de su sexagésimo aniversario* (pp. 629-644, vol. 2). Arco Libros.
- FREIXEIRO MATO, X. R. (2006). *Gramática da lingua galega. II. Morfosintaxe* (2.ª ed.). A Nosa Terra.
- GIVÓN, T. (2001). *Syntax: An introduction* (2.ª ed.). John Benjamins. <https://doi.org/10.1075/z.syns>
- GONZÁLEZ-VÁZQUEZ, M. (2021). The evidentiality system in Galician and the *seica* marker. *Journal of Pragmatics*, 178, 83-92. <https://doi.org/10.1016/j.pragma.2021.03.009>
- GREENACRE, M. (2007). *Correspondence analysis in practice* (2.ª ed.). Chapman & Hall; CRC. <https://doi.org/10.1201/9781420011234>
- HALLIDAY, M. A. K. (2014). *Halliday's introduction to Functional Grammar* (4.ª ed.). Routledge. <https://doi.org/10.4324/9780203431269>
- HOYE, L. (1997). *Adverbs and modality in English*. Longman.
- HU, C. e LI, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, 8(6), 20-31. <https://doi.org/10.5539/elt.v8n6p20>
- LÊ, S., JOSSE, J. e HUSSON, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1-18. <https://doi.org/10.18637/jss.v025.i01>
- LEVSHINA, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins. <https://doi.org/10.1075/z.195>
- LUNA, M., VILLALÓN, R., MATEOS, M. e MARTÍN, E. (2020). Improving university argumentative writing through online training. *Journal of Writing Research*, 12(1), 233-262. <https://doi.org/10.17239/jowr-2020.12.01.08>
- MCENERY, T. e KIFLE, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. En J. Flowerdew (Ed.), *Academic discourse* (pp. 182-195). Routledge.
- MÍGUEZ REGO, V. (2021). *Aproximación á modalidade epistémica en galego: o dominio adverbial* [Tese de doutoramento, Universidade de Santiago de Compostela]. Minerva. <http://hdl.handle.net/10347/27126>
- MÍGUEZ, V. (2021a). A categoría da modalidade e a súa expresión na lingua galega.

Estudos de Lingüística Galega, 13, 107-137. <https://doi.org/10.15304/elg.13.7386>

- MÍGUEZ, V. (2021b). The diachrony of Galician *certamente* and *seguramente*: A case of grammatical constructionalization. En M. Hilpert, B. Cappelle e I. Depraetere (Eds.), *Modality and diachronic construction grammar* (pp. 123-148). John Benjamins. <https://doi.org/10.1075/cal.32.05mig>
- MÍGUEZ, V. (2022). On epistemic modality and discourse strategy: Evidence from Galician adverbs. *Journal of Pragmatics*, 201, 32-42. <https://doi.org/10.1016/j.pragma.2022.09.003>
- NUYTS, J. (2001). *Epistemic modality, language, and conceptualization: A cognitive-pragmatic perspective*. John Benjamins. <https://doi.org/10.1075/hcp.5>
- PALMER, F. R. (1990). *Modality and the English modals* (2.^a ed.). Longman.
- PIETRANDREA, P. (2008). *Certamente* and *sicuramente*: Encoding dynamic and discursive aspects of commitment in Italian. *Belgian Journal of Linguistics*, 22(1), 221-246. <https://doi.org/10.1075/bjl.22.11pie>
- R CORE TEAM (2022). *R: a language and environment for statistical computing* [versión 4.2.1]. R Foundation for Statistical Computing. <https://www.R-project.org>
- RODRÍGUEZ-ESPIÑEIRA, M. J. (2010). Modalidad, gramática y discurso: *posible, probable* y sus antónimos. En M. J. Rodríguez-Espiñeira (Ed.), *Adjetivos en discurso. Emociones, certezas, posibilidades y evidencias* (pp. 181-253). Universidade de Santiago de Compostela.
- YANG, D. E ZENG, X. (2021). Epistemic modality and projection in argumentative writing: A corpus-driven comparison of Chinese English majors and native English speakers. *Theory and Practice in Language Studies*, 11(4), 412-421. <https://doi.org/10.17507/tpls.1104.11>

APÉNDICE

TÁBOA 4. Lemas segundo o tipo expresivo

Tipo de expresión	Lema	#Tokens	Contexto
Adverbio	ao mellor	13	Os novos pratos conteñen <i>ao mellor</i> trazos de cociña de sempre, mais cambiando por completo a súa estética, gornicións, presentación...
	igual	3	Na miña opinión os valores da sociedade de hoxe en día deberían cambiar moito, fixarse na persoa e non na roupa, non estar influenciados por persoas ou pola publicidade, ter criterio propio e en vez de gastarse diñeiro consumindo e consumindo gardáramolo para cousas máis importantes, <i>igual</i> a vida non sería tan desigual, nin habría tantos niveis, xa que tamén o nivel de sociedade dunha persoa depende do que consuma e canto consuma
	posiblemente	24	As condicións de traballo rozan os límites pero os consumidores non pensan neso, só no precio (egoísmo) e <i>posiblemente</i> moitos traballadores ou traballadoras compren roupa alí sen prever que se están autoperxudicando
	probablemente	34	Os programas didácticos teñen sen moitas variacións a mesma audiencia que hai dez anos. Estes tratan sobre recetas sinxelas e cuxa preparación na casa é plausible. Teñen verdadeira utilidade xa que <i>probablemente</i> calquera é capaz de facer un dos seus pratos porque mostran a preparación paso a paso
	quizá	7	En conclusión, <i>quizá</i> a cultura galega en certos ámbitos non sexa o máis interesante para os rapaces, pero non é por que soamente soñen con modelos e deportistas, senon porque buscan outro tipo de cultura sexa artística, científica ou doutro tipo...
	quizais	50	O máis raro é que <i>quizais</i> país que viron cómo os seus propios país vivían felices con pouco e incluso puido faltarlle de todo, a día de hoxe crían aos seus fillos baixo o capitalismo puro e duro; os nenos son bombardeados con agasallos, desexos, obxetos inútiles, todo canto piden...
	quizás	10	Un sector do que non paramos nunca de formar parte é o da alimentación, pois sen comer non podemos vivir, e <i>quizás</i> sexa un do que mais se aproveiten
	se cadra	2	Como non, as grandes empresas como «Apple», «Adidas» ou «Nike», aproveítanse desta mentalidade sacando teléfonos móbiles de última xeneración que <i>se cadra</i> teñen menor durabilidade que outra marca menos prestixiosa, pero hoxe en día iso é o que importa, gastarte cartos en móbiles e roupa de marca
	seguramente	29	Este cambio <i>seguramente</i> ven dado pola demanda, maioritariamente de clases medias e altas, de probaren novas experiencias, neste caso, gustativas
	seica	1	Así pois, aínda que sabemos que unha leituga do mercado ou calquer outro vexetal é máis sano, máis ecolóxico e máis natural, compramos a leituga do Eroski que é máis grande e sobre todo máis barata e así, cos cartos que aforramos, mercamos un novo móbil, outro coche e, <i>seica</i> , se nos atopamos nunha época de «BOOM» inmobiliario nos metemos nunha hipoteca dun piso adquirido mediante un préstamo que non seremos capaces de devolver
talvez	15	Isto que estamos a facer, uns 20 anos cara atrás non ocorría, tiñas que aguantar cos mesmos zapatos meses e ningún se queixaba; <i>tal vez</i> iso pasaba porque non se tiña diñeiro suficiente pero agora as cousas tampouco están como para tirar a casa pola fiestra	

TÁBOA 4. Continuación

Tipo de expresión	Lema	#Tokens	Contexto
adxectivo	posible	13	Estes avances están ben, xa que indican que a sociedade evoluciona, pero sempre e cando, non se leven a situacións de conflito ou non éticas. Como por exemplo, a diferenza de prezos e a cantidade de comida entre lugares máis prestixiosos ou menos. <i>É posible</i> que dentro duns anos o mundo gastronómico perda importancia se desaparecen todos eses programas televisivos, pero aínda así seguirá evolucionando coa tecnoloxía, xa que novos obxectos de cociña e formas de cociñar aparecerán co tempo e tamén novos alimentos serán introducidos, xa sexa pola aparición ou polo descubrimento neste ámbito
	posíbel	1	<i>É posíbel</i> que a cociña tal e como a coñecemos desapareza pero haberá novas correntes onde se empregarán novos produtos para obter novos sabores
	probable	21	O mundo da gastronomía, igual que o mundo da moda está en constante cambio e renovación, <i>é probable</i> que os novos cambios deixen de agradar ao público, que os programas deixen de ter interese, e que as persoas dense conta que, polo prezo que pagas esa «delicia» comería perfectamente unha familia no tercer mundo
auxiliar	poder	174	Como conclusión, é necesario que a xente baixe o nivel de consumo, xa que moi a curto prazo isto <i>podérase</i> volver perigoso
verbo	coídar	2	<i>Coído</i> que sería interesante que os expertos en economía se puxeran máns a obra para atopar alternativas para unha sociedade na que se priman os restaurantes antes que a saúde ou a moral
	crer	79	Do mesmo xeito apareceron os gastrónomos porque <i>creo</i> que precisarían xente para catar a comida e proclamar quen é o mellor cocineiro
	pensar	99	<i>Penso</i> que tamén inflúen moitas veces nese desexo de carto rápido a nosa propia familia, que polo menos nalgúns casos, incitan aos seus fillos a xogar ao fútbol cunha presión innecesaria e llo transmiten aos seus fillos

TÁBOA 5. Resumo dos resultados

		Adverbio	Adxectivo	Auxiliar	Verbo
(inter)subxectividade	subxectivo	131	2	83	179
	intersubxectivo	57	33	91	0
performatividade	performativo	188	35	174	167
	descriptivo	0	0	0	12
focalidade	non-foco	188	35	174	179
	foco	0	0	0	0
tema	non-tema	139	22	150	78
	tema	49	13	24	101
complementación	non-compl.	188	0	147	2
	complementador	0	35	27	177

TÁBOA 5. Continuación

polaridade	afirmativa	152	29	173	137
	negativa	36	6	1	42
información secundaria	non-secundaria	143	30	152	171
	secundaria	45	5	22	8
modificadores	non-modificador	185	17	174	179
	modificador	3	18	0	0
estratexia discursiva	non-estratéxico	179	32	120	174
	atenuación	5	1	24	2
	concesión	4	2	30	3

TÁBOA 6. Exemplos de expresións epistémicas modificadas (CORTEGAL)

Expresión	#Tokens	Contexto
moi probablemente	2	O diñeiro move o mundo e o mundo quere ter máis, producir máis, polo que <i>moi probablemente</i> , as ventas e as compras aumentarán no futuro cercano
moi posiblemente	1	Polo xeneral amólanos todo, somos moi irascibles. Pero por outra banda, creo que nosos pais deberían entendernos, eles <i>moi posiblemente</i> pasaron por algo similar, aínda que fosen outros tempos
moi probable	6	Na actualidade todo o producido ten a finalidade de ser consumido, polo que se esta venta ten éxito no futuro <i>é moi probable</i> que a produción dese obxecto-servizo sexa explotada ata a saciedade
pouco probable	1	Tamén dase o caso de non sexa simplemente unha moda e definitivamente estableza un cambio cultural pero na miña opinión o vexo inviable xa que na sociedade na que vivimos infestada pola comida lixo <i>vexo pouco probable</i> que o conxunto da sociedade cambie xa que a opción de que sexa unha simple moda é máis razoable e comprensible
moi pouco probable	1	A algunhas persoas pareceralles raro pero, si unha persoa non viaxa, <i>é moi pouco probable</i> que tome comida á que non está acostumada, polo que se está perdendo parte das costumes doutros países
máis probable	7	As causas deste auxe poden deberse entre outras aos turistas, é dicir, cando un grupo de xente viaxa a un determinado país e ese país gustoulle pola súa comida <i>o máis probable</i> e que regrese a el e que difunda isto entre os seus coñecidos
moi posible	2	No día a día non nos percatamos pero <i>é moi posible</i> que si miramos cara uns anos atrás, comprobemos cómo cambiaron as cousas
máis posible	1	Aínda cando creemos que mercamos algo por necesidade, probablemente non sexa así, <i>o máis posible</i> é que esteamos sendo manipulados pola industria capitalista

TÁBOA 7. Contribución das variables á Dimensión 1 (ACM)

	R²	Valor p
complementación	0,54114521	3,595869e-99
tema	0,48145032	6,684181e-84
(inter)subxectividade	0,36410399	2,052041e-58
información secundaria	0,23982974	4,481322e-36
estratexia	0,19277847	1,526088e-28
negación	0,05275016	2,476767e-08
performatividade	0,02732380	6,719262e-05

TÁBOA 8. Contribución das variables á Dimensión 2 (ACM)

	R²	Valor p
modificadores	0,58709289	2,438042e-112
(inter)subxectividade	0,29552049	1,328749e-45
complementación	0,13822730	2,541744e-20
tema	0,07403944	3,094858e-11
estratexia	0,06126255	1,729791e-09
negación	0,04514216	2,646761e-07
información secundaria	0,04069146	1,055187e-06
performatividade	0,01038009	1,443540e-02

ÍNDICE DE TÁBOAS

***CORTEGAL*, Corpus de textos galegos escritos por estudantes no ámbito académico. Deseño do corpus e caracterización dos textos**

- 22 TÁBOA 1. Distribución de exames por convocatoria e temática na mostra de *CORTEGAL*
- 23 TÁBOA 2. Notas numéricas da mostra de *CORTEGAL*
- 50 TÁBOA 3. Distribución dos códigos descritivos das formas non estándar por nivel lingüístico
- 55 TÁBOA 4. Códigos empregados para explicar a orixe da forma diverxente
- 73 TÁBOA 5. Códigos empregados para a anotación dos conectores que vinculan enunciados
- 77 TÁBOA 6. Datos cuantitativos sobre os textos de *CORTEGAL*
- 78 TÁBOA 7. Distribución dos *tokens* en clases de palabras
- 80 TÁBOA 8. Códigos descritivos das formas non estándar por nivel
- 81 TÁBOA 9. Códigos descritivos con máis de 500 ocorrencias
- 82 TÁBOA 10. Distribución dos códigos de orixe
- 84 TÁBOA 11. Distribución dos códigos dos conectores que vinculan enunciados

O léxico non estándar en *CORTEGAL*

- 120 TÁBOA 1. Distribución das formas non estándar do nivel léxico segundo o aspecto diverxente do estándar
- 121 TÁBOA 2. Distribución das formas non estándar do nivel léxico de acordo coa súa orixe
- 122 TÁBOA 3. Distribución por categorías gramaticais das formas léxicas non estándar (I)
- 123 TÁBOA 4. Distribución por categorías gramaticais das formas léxicas non estándar (II)
- 128 TÁBOA 5. Distribución das formas non estándar con diferenzas que van máis alá do xénero e da acentuación en función da súa orixe
- 129 TÁBOA 6. Nivel de presenza das formas non estándar con cinco ou máis ocorrencias en relación coas correspondentes formas estándar
- 130 TÁBOA 7. Formas non estándar cun nivel de presenza igual ou superior ao das formas estándar correspondentes
- 134 TÁBOA 8. Distribución por categorías gramaticais das transferencias directas do español (con diferenzas que van máis alá do xénero e a acentuación)
- 135 TÁBOA 9. Calcos estruturais de expresións pluriverbais fixas do español
- 137 TÁBOA 10. Calcos estruturais de compostos léxicos e sintagmas univocalizados do español
- 138 TÁBOA 11. Calcos estruturais de palabras derivadas e outras palabras complexas
- 139 TÁBOA 12. Adaptacións que afectan a elementos consonánticos
- 141 TÁBOA 13. Adaptacións que afectan á acentuación
- 142 TÁBOA 14. Adaptacións que afectan a ditongos
- 143 TÁBOA 15. Adaptacións que afectan a terminacións
- 145 TÁBOA 16. Combinacións de galego e castelán sen padrón regular
- 147 TÁBOA 17. Modificacións que afectan a elementos consonánticos

- 148 TÁBOA 18. Modificacións que afectan a ditongos
- 149 TÁBOA 19. Modificacións que afectan a terminacións
- 152 TÁBOA 20. Formas creadas por analogía con outras voces da familia de palabras ou de familias formalmente próximas
- 154 TÁBOA 21. Formas creadas por analogía con outras voces do paradigma flexivo
- 156 TÁBOA 22. Formas creadas con alomorfos dos sufixos estándar
- 157 TÁBOA 23. Formas creadas mediante sufixación con competidores coa mesma raíz
- 158 TÁBOA 24. Formas creadas mediante derivación sen competidores coa mesma raíz
- 158 TÁBOA 25. Palabras cuxa forma non se corresponde co significado buscado
- 159 TÁBOA 26. Combinacións «imposibles» de bases e afixos
- 160 TÁBOA 27. Voces cuxa forma se pode explicar pola recorrencia de certas terminacións
- 162 TÁBOA 28. Voces non estándar de orixe galega presentes no galego popular
- 164 TÁBOA 29. Formas transferidas desde o galego elaborado
- 165 TÁBOA 30. Estranxeirismos non adaptados
- 166 TÁBOA 31. Formas resultantes da mestura de palabras e expresións
- 167 TÁBOA 32. Formas non estándar con orixe de difícil clasificación
- 169 TÁBOA 33. Formas con asignación inadecuada de xénero por influencia do castelán
- 171 TÁBOA 34. Formas con asignación inadecuada de xénero por analogía
- 172 TÁBOA 35. Formas cuxa asignación inadecuada de xénero non se explica nin por transferencia nin por analogía
- 174 TÁBOA 36. Formas con cambio na posición do acento de intensidade por probable influencia do español
- 175 TÁBOA 37. Formas con cambio na posición do acento de intensidade non explicables por influencia do español
- 187 TÁBOA 38. Presenza no *TILG*, *CORGA* e *TLPGP* das formas non estándar cun nivel de aparición en *CORTEGAL* igual ou superior ao das formas estándar correspondentes así como das súas variantes
- 189 TÁBOA 39. Presenza no *TILG*, *CORGA* e *TLPGP* de calcos estruturais do español
- 190 TÁBOA 40. Presenza no *TILG*, *CORGA* e *TLPGP* de formas adaptadas do español
- 194 TÁBOA 41. Presenza no *TILG*, *CORGA* e *TLPGP* de combinacións de galego e castelán sen padrón regular
- 195 TÁBOA 42. Presenza no *TILG*, *CORGA* e *TLPGP* de hipergaleguismos
- 197 TÁBOA 43. Presenza no *TILG*, *CORGA* e *TLPGP* de voces creadas por analogía con outras formas galegas
- 199 TÁBOA 44. Presenza no *TILG*, *CORGA* e *TLPGP* de voces non estándar presentes no galego popular
- 200 TÁBOA 45. Presenza no *TILG*, *CORGA* e *TLPGP* de voces non normativas presentes no galego elaborado
- 200 TÁBOA 46. Presenza no *TILG*, *CORGA* e *TLPGP* de estranxeirismos non adaptados
- 200 TÁBOA 47. Presenza no *TILG*, *CORGA* e *TLPGP* de unidades que combinan elementos de varias palabras ou expresións galegas

- 201 TÁBOA 48. Presenza no *TILG*, *CORGA* e *TLPGP* de unidades léxicas non estándar con diferenzas que van máis alá do xénero e da acentuación con respecto ao estándar cuxa orixe non é atribuíble ás razóns sinaladas nas táboas previas
- 201 TÁBOA 49. Presenza no *TILG*, *CORGA* e *TLPGP* de palabras con xénero non estándar
- 202 TÁBOA 50. Presenza no *TILG*, *CORGA* e *TLPGP* de palabras con acentuación non estándar

A posición do pronome persoal átono con formas verbais finitas no corpus *CORTEGAL*

- 209 TÁBOA 1. Casos totais de pronomes átonos analizados con respecto ó total de *CORTEGAL*
- 210 TÁBOA 2. Casos totais de énclice, de próclise e de énclice e próclise á vez con formas verbais finitas
- 211 TÁBOA 3. Datos de énclice e próclise no corpus *CORTEGAL* en comparación cos *Contos populares*
- 212 TÁBOA 4. Énclice e próclise segundo a persoa pronominal
- 212 TÁBOA 5. Grao de coincidencia da posición do clítico en *CORTEGAL* con respecto ó galego estándar escrito
- 220 TÁBOA 6. Indutores de próclise canónica en *CORTEGAL*
- 221 TÁBOA 7. Lista de proclisadores nexuais en *CORTEGAL*
- 222 TÁBOA 8. Estructuras sintácticas en que se deu próclise nexual
- 224 TÁBOA 9. Secuencias de nexos e clítico en próclise en subordinadas
- 226 TÁBOA 10. Constituíntes funcionais entre o nexos e o clítico
- 227 TÁBOA 11. Palabras de negación que provocan próclise
- 228 TÁBOA 12. Adverbios focalizados que provocan próclise
- 230 TÁBOA 13. Cuantificadores focalizados que provocan próclise
- 234 TÁBOA 14. Casos de colocacións anómalas de énclice por próclise
- 236 TÁBOA 15. Tipoloxía dos nexos que non provocaron próclise
- 242 TÁBOA 16. Tipoloxía de oracións en que se dan casos de énclice pronominal cando hai un constituínte entre o nexos e o verbo
- 244 TÁBOA 17. Funcións sintácticas dos constituíntes inseridos entre o nexos e o verbo en casos de énclice pronominal
- 245 TÁBOA 18. Ocorrencias de nexos que non provocan próclise
- 246 TÁBOA 19. Próclise e énclice en oracións subordinadas con nexos
- 247 TÁBOA 20. Porcentaxes de próclise e énclice nas subordinadas con nexos máis numerosas

A expresión da incerteza nos textos de *CORTEGAL*: un achegamento paradigmático

- 326 TÁBOA 1. Factores decisivos no uso das expresións epistémicas (Nuyts, 2001, p. 227)
- 328 TÁBOA 2. Padróns sintácticos das expresións epistémicas (Nuyts, 2001, p. 262)
- 331 TÁBOA 3. Distribución de *tokens* e lemas por tipo de expresión
- 347 TÁBOA 4. Lemas segundo o tipo expresivo
- 348 TÁBOA 5. Resumo dos resultados

- 349 TÁBOA 6. Exemplos de expresións epistémicas modificadas (*CORTEGAL*)
- 350 TÁBOA 7. Contribución das variables á Dimensión 1 (ACM)
- 350 TÁBOA 8. Contribución das variables á Dimensión 2 (ACM)

ÍNDICE DE GRÁFICAS

CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico. Deseño do corpus e caracterización dos textos

- 23 GRÁFICA 1. Cualificacións dos textos de *CORTEGAL*
- 24 GRÁFICA 2. Distribución das cualificacións numéricas nos textos de *CORTEGAL*
- 24 GRÁFICA 3. Distribución das cualificacións numéricas do texto en xuño e setembro
- 26 GRÁFICAS 4,5. Lingua inicial e lingua habitual da mocidade galega (15-29 anos) (Fonte: IGE, 2019)
- 80 GRÁFICA 6. Distribución das formas non estándar nos textos de *CORTEGAL*
- 84 GRÁFICA 7. Distribución dos conectores nos textos de *CORTEGAL*

Análise da acentuación gráfica a través do corpus *CORTEGAL*

- 97 GRÁFICA 1. Clasificación dos erros de omisión
- 98 GRÁFICA 2. Clasificación dos erros de adición
- 100 GRÁFICA 3. Clasificación dos erros de desprazamento
- 101 GRÁFICA 4. Clasificación dos erros de acentuación segundo as diferenzas de acentuación entre o galego e o castelán
- 103 GRÁFICA 5. Distribución dos erros segundo a estrutura acentual da palabra afectada
- 103 GRÁFICA 6. Incidencia dos erros de colocación segundo a estrutura acentual
- 104 GRÁFICA 7. Incidencia dos erros segundo a estrutura acentual da palabra ordenados polo tipo de fenómeno de colocación do acento
- 108 GRÁFICA 8. Dispersión e frecuencia dos erros de acentuación por exame respecto ao total de exames de *CORTEGAL*
- 109 GRÁFICA 9. Dispersión e frecuencia dos erros de omisión por exame
- 110 GRÁFICA 10. Distribución e frecuencia dos erros de adición por exame

O léxico non estándar en *CORTEGAL*

- 119 GRÁFICA 1. Correlación entre o número de formas non estándar do nivel léxico e o número de textos
- 120 GRÁFICA 2. Correlación entre o número de ocorrencias das formas non estándar do nivel léxico e o número de formas non estándar
- 124 GRÁFICA 3. Comparativa entre a distribución das categorías gramaticais no total do corpus e nas formas léxicas non estándar

Procedementos de xeneralización e marcado da implicación no corpus *CORTEGAL*: entre a linguaxe académica e o rexistro coloquial da lingua

- 303 GRÁFICA 1. Total de referencias explícitas, mediante o pronome persoal ou o posesivo, á persoa emisora do discurso

ÍNDICE DE FIGURAS

CORTEGAL, Corpus de textos galegos escritos por estudantes no ámbito académico. Deseño do corpus e caracterización dos textos

- 29 FIGURA 1. Visualización das anotacións lingüísticas dos *tokens* (I)
- 30 FIGURA 2. Visualización das anotacións lingüísticas dos *tokens* (II)
- 31 FIGURA 3. Visualización do texto na súa forma orixinal
- 32 FIGURA 4. Visualización do texto sen a disposición de liñas orixinal e sen distribución en páxinas
- 32 FIGURA 5. Visualización do texto con marcas de inicio e remate de cada liña e páxina
- 33 FIGURA 6. Visualización das formas riscadas
- 33 FIGURA 7. Visualización da versión final da/do estudante nun texto con formas riscadas
- 34 FIGURA 8. Visualización da palabra engadida *mesmo*
- 34 FIGURA 9. Visualización da palabra de lectura dubidosa *tamén*
- 36 FIGURA 10. Fragmento do formulario para a introdución de metadatos
- 37 FIGURA 11. Metadatos visibles para as persoas usuarias do corpus
- 40 FIGURA 12. Anotación e normalización da forma *durara* nun texto de *CORTEGAL*
- 44 FIGURA 13. Normalización e codificación en TEITOK
- 45 FIGURA 14. Dobre estandarización e conservación das modificacións en TEITOK
- 46 FIGURA 15. Dobre estandarización sen conservación das modificacións en TEITOK
- 48 FIGURA 16. Visualización dun texto na capa de estandarización morfolóxica
- 48 FIGURA 17. Visualización dun texto na capa de estandarización discursiva
- 51 FIGURA 18. Asignación de varios códigos de forma non estándar a unha mesma palabra
- 53 FIGURA 19. Ligazón para consultar as anotacións multipalabra
- 53 FIGURA 20. Pantalla que se visualiza tras premer na ligazón Anotación multipalabra
- 58 FIGURA 21. Anotación dunha corrección derivada
- 59 FIGURA 22. Visualización das formas con corrección derivada
- 61 FIGURA 23. Visualización dunha corrección derivada nun caso de concordancia
- 70 FIGURA 24. Frecuencia de D_reg_su por lema orixinal (fragmento da táboa xerada por TEITOK)
- 71 FIGURA 25. Frecuencia de L_w_su por lema orixinal (fragmento da táboa xerada por TEITOK)
- 71 FIGURA 26. Frecuencia de L_w_su por lema estándar (fragmento da táboa xerada por TEITOK)

Análise da acentuación gráfica a través do corpus *CORTEGAL*

- 106 FIGURA 1. Frecuencia de aparición dos padróns acentuais do español. Fonte: Quilis (1983)

A expresión da incerteza nos textos de *CORTEGAL*: un achegamento paradigmático

- 340 FIGURA 1. Mapa da Análise de Correspondencia Múltiple (ACM)

