

Contribuíndo ao avance das Humanidades Dixitais: Recursos para o estudo do inglés histórico e contemporáneo

II Xeira
CLARIAH-GAL
09/05/2025



María José López Couso (Coordinadora), Zeltia Blanco Suárez, Tamara Bouso, Teresa Fanego, Fátima Faya, Belén Méndez Naya, Paloma Núñez Pertejo e Paula Rodríguez Abruñeiras



O grupo *Variación, Cambio Lingüístico e Gramaticalización* (VLCG), fundado en 1991 por **Teresa Fanego**, centra a súa atención no estudo da **variación e o cambio lingüístico** nos niveis morfosintáctico, semántico e pragmático-discursivo dende os puntos de vista diacrónico, diatópico e contrastivo. Na súa actividade investigadora, o grupo VLCG adopta unha perspectiva eminentemente variacionista, empregando a metodoloxía da **lingüística de corpus** e a aproximación teórica dos estudos de **gramaticalización**, a **Lingüística Cognitiva** e a **Gramática de Construcións**. Os membros do grupo VLCG contan, ademais, cunha dilatada traxectoria na compilación de **corpus computarizados**.

1 CHELAR (*Corpus of Historical English Law Reports 1535-1999*)

CHELAR é un corpus de documentos legais ingleses compilado polo grupo VLCG entre 2011 e 2018. En abril de 2016 completouse a primeira versión do corpus (**CHELAR v.1**), en texto plano e con etiquetaxe morfolóxica. En decembro de 2018 saíu á luz a segunda versión (**CHELAR v.2**; ver **Figura 1**), que incorpora **etiquetaxe XML** (ver **Figura 2**). CHELAR ten **múltiples aplicacións** nos ámbitos da lingüística diacrónica e o estudo do discurso especializado; responde tamén ao crecente interese das institucións europeas pola linguaxe legal e parlamentaria. Para poder utilizar este recurso, é necesario **rexistrarse**; as instrucións pódense atopar no código QR.



2 ARCHER (*A Representative Corpus of Historical English Registers*)

Compilado por un **consorcio internacional** de 14 universidades europeas e norteamericanas, ARCHER fai posible a análise sistemática dunha variedade de rexistros textuais e a súa evolución no período 1600-1999. As distintas versións de ARCHER publicadas ata o momento contan con **etiquetaxe gramatical** (ver **Figura 3 A**) e **XML**. Na súa última versión (ARCHER 3.2), o corpus contén 3,3 millóns de palabras de inglés británico e americano. O grupo VLCG foi o responsable da compilación dos **textos legais** do corpus (ver **Figura 3 B**). Dende a súa aparición a principios dos anos 90, o corpus ten sido empregado como base para **máis de 1.000 publicacións**, o que o converte nun recurso esencial para especialistas en lingüística inglesa a nivel mundial. Para facer uso deste recurso, é necesario **rexistrarse**; as instrucións pódense atopar no apartado “Using Archer”, no código QR.



Original texts (txt and PDF)			
RAW	POS-annotated	TEI-XML-annotated	
Subperiod	Words	Files	Texts
1950-1999	50,543	20	20
1900-1949	50,728	20	21
1850-1899	51,320	20	24
1800-1849	52,182	20	32
1750-1799	51,498	20	21
1700-1749	50,794	20	22
1650-1699	50,653	20	30
1600-1649	51,050	22	112
1535-1599	54,558	23	87
Total	463,326	185	369

Figura 1. Estrutura do CHELAR v.2

```
<lb/>AFFRAN from <name type="person">Brandon J.</name>
<lb/>By a writ dated <date when="18-06-1968">June 18, 1968</
date>, the plaintiffs, La Plata Cereal Co. S.A.,
<lb/>of <place>Buenos Aires</place> ("shippers"), André S&mp;
Cie. S.A. ("Swiss André"), of
<lb/><place>Lausanne</place>, Comptoir Commercial André Cie.,
of <place>Marseille</place> ("French
<lb/>André") and Sorveglia, of <place>Rome</place>
(collectively described as owners of
<lb/>the <foreign lang="sp">cargo</foreign> lately laden on
board the vessel the <hi rend="italics">Annefeld</hi>, began an
action
<lb/><foreign lang="la">in rem</foreign> against the
vessel owned by a Liberian company, Asmarfield <choice
n="hyphenation"><orig>Ship<lb break="no">/>ping</
orig><reg>Shipping
<lb/></reg></choice> Corporation ("the shipowners") for damages
for breach of contract
[...]
```

Figura 2. Etiquetaxe XML no CHELAR v.2

(A)
That DD1 is VBZ the AT transaction_NN1 in II which DDQ the AT trustees_NN2 here RL made VVD a AT1 disposition_NN1 ((the AT grant_NN1 to II Mr. NNB Roberts_NP1 of IO new JJ rights_NN2 over II the AT pictures_NN2)) , , and CC the AT transaction_NN1 must VBI be VBI tested VVN by II asking_VVG : : ((I MC1))) was_VBDZ that DD1 disposition_NN1 ((the AT grant_NN1 of IO the AT new JJ rights_NN2)) intended_VVD to TO confer_VVI a AT1 gratuitous JJ benefit_NN1 ?
(B)
reasons) ; in such a case the transaction in which the disposition is made will inevitably include the associated operations that effect it . exactly how section 20(4) should operate in every possible case of a disposition by an individual . Where paragraph 6(3) applies , the court does That is the transaction in which the trustees here made a disposition (the grant to Mr. Roberts of new rights over the pictures transaction must be tested by asking : (i) was that disposition (the grant of the new rights) intended to confer a a transaction falling to be tested under section 20(4) may include a disposition effected by associated operations . Section 20(4) is a very important part The commencement of the interest in possession under the settlor's original disposition would still have been an occasion of charge but could hardly be power of appointment they only give effect to the settlor's original disposition , as it were by filling in a blank ; they confer are directed to taxing the value transferred by a voluntary , bountiful disposition , and death is treated as the occasion of such a disposition disposition , and death is treated as the occasion of such a disposition . A good deal of adaptation is required to apply this basic

Figura 3. Exemplo de etiquetaxe gramatical (A) e concordancia no ARCHER (B)

3 COLMOBAENG (*A Corpus of Late Modern British & American English Prose*)

O corpus, de 1.170.000 palabras, inclúe textos de ficción e non ficción dos **séculos XVIII e XIX**, tanto de inglés británico coma de inglés americano.

5 TTT (*Toledo Teachers Trainees Corpus*)

Trátase dun corpus de 170.000 palabras que reúne datos de 354 estudantes universitarias/os de Formación do Profesorado sobre a súa **percepción** sobre o proceso de **aprendizaxe de linguas** (profesorado, materiais, contidos, metodoloxías e avaliación).

6 PELEC (*Primary Education Learners' English Corpus*)

PELEC é un corpus de aprendices de inglés con datos de 252 estudantes de 4º e 6º de Educación Primaria de centros públicos na Comunidade Autónoma de Cantabria, tanto de colexios adheridos a programas de **Aprendizaxe Integrada de Contidos en Lingua Estranxeira** (AICLE) coma de centros que non ofertan programas AICLE. O corpus contén **datos orais e escritos** que suman un total de 61.609 palabras.

Referencias
Blanco Suárez, Zeltia, Francisco Gallardo del Puerto e Evelyn Gandón Chapela. 2020. The Primary Education Learners' English Corpus (PELEC): Design and compilation. *Research in Corpus Linguistics* 8(1): 147-163.
BSNC = Bouso, Tamara e Pablo Ruano San Segundo. 2024. compiladores. *British Sentimental Novel Corpus*. ISBN 978-84-09-60357-2.
Fanego, Teresa. 2012. COLMOBAENG: A Corpus of Late Modern British and American English Prose. En Nila Vázquez (ed.), *Creation and use of historical English corpora in Spain*, 101-117. Cambridge Scholars.
Fanego, Teresa, Paula Rodríguez Puente, María José López Couso, Belén Méndez Naya, Paloma Núñez Pertejo, Cristina Blanco García e Iván Tamaredo. 2017. The Corpus of Historical English Law Reports 1535-1999 (CHELAR): A resource for analysing the development of English legal discourse. *ICAME Journal* 41: 53-82.
Faya, Fátima e Gema Alcaraz Mármol. 2020. The Toledo Teachers Trainees (TTT) Corpus: Bridging the gap between students' narratives and corpus linguistics. *Research in Corpus Linguistics* 8(1): 164-177.
López Couso, María José e Belén Méndez Naya. 2012. Compiling British English legal texts: A contribution to ARCHER. En Nila Vázquez González (ed.), *Creation and use of historical English corpora in Spain*, 5-19. Cambridge Scholars.

4 BSNC (*British Sentimental Novel Corpus*)

Contén **21 millóns de palabras** extraídas de **114 novelas completas** de 11 novelistas do canon do **século XIX**, entre outros Jane Austen, Mary Shelley, Walter Scott e Charles Dickens (ver **Figura 4**). De **libre acceso** (ver código QR).



Figura 4. Novelistas incluídos no BSNC

Pola súa amplitude temática e a súa extensa cobertura temporal e textual, os corpus compilados nos últimos anos polo grupo VLCG representan unha contribución relevante ao desenvolvemento das Humanidades Dixitais e proporcionanlle á comunidade científica unha ferramenta de enorme potencial que permite obter un maior e mellor coñecemento da lingua inglesa histórica e contemporánea.

FINANCIAMENTO: ED431B 2023/03; PID2023-146887NB-I00.