

Que?	Por que?
<ul style="list-style-type: none">- Crear ferramentas de análise automática de poesía e narrativa en galego:<ul style="list-style-type: none">- escansión automática- detección de rima e encabalgamento- clasificación de estrofa- Comparación multilingüe dos seus textos con obras en portugués e castelán<ul style="list-style-type: none">- adaptación de modelos de linguaxe do proxecto Nós (Gamallo et al., 2024) e anteriores (García, 2021)- Distribución das ferramentas con licenza libre	<ul style="list-style-type: none">- Falta de recursos, que xa existen nas linguas europeas maioritarias, para a análise computacional de literaturas en linguas minoritarias<ul style="list-style-type: none">- cf. poesía en castelán: Agirrezabal (2017), de la Rosa et al. (2020; 2021), Marco & Gonzalo (2021), Navarro (2018), Pérez et al. (2022), Ruiz et al. (2017)- cf. poesía en portugués: Araújo & Mamede (2002); Mittmann (2016), Mittmann et al. (2019)
Corpus	Desafíos
<ul style="list-style-type: none">- Rexurdimento, Prerrexurdimento, período intersecular, até a fin do dominio público- Escolla segundo obras recoñecidas na historia da literatura (p. ex. Carballo Calero, 1975; Vilavedra, 1999; Pena, 2014)- Prefírense fontes en formato electrónico cando están dispoñíbeis (Galiciiana)	<ul style="list-style-type: none">- Falta de texto electrónico público fiábel (a pesar da dispoñibilidade de dixitalizacións)- Variación scriptolingüística nas obras históricas que complica as comparacións de contido

Exemplo das anotacións a automatizar (expresadas en formato TEI)

```
<lg type="quintilla">
  <l met="--+--+--" rhyme="a">Mais son probe, e mal <w type="rhyme">pecado</w>,</l>
  <l met="--+--+--" rhyme="b">A miña terra n' é <w type="rhyme">miña</w>,</l>
  <l met="+--+--+" rhyme="a" enjamb="ex_obj_verb">Qu' hastra lle dán de <w type="rhyme">prestado</w></l>
  <l met="+-----" rhyme="b">A veira por que <w type="rhyme">camiña</w></l>
  <l met="---+---" rhyme="a">Ô que naceu <w type="rhyme">desdichado</w>.</l>
</lg>
```

Referencias

Agirrezabal. (2017). *Automatic Scansion of Poetry*. [Tese de doutoramento. University of the Basque Country]. Carballo Calero, R. (1975) [1963]. *Historia da Literatura Galega Contemporánea*. Galaxia.. De la Rosa, J. et al. (2021). Transformers analyzing poetry: Multilingual metrical pattern prediction with transformer-based language models. *Neural Computing and Applications*. Araújo, P. A., & Mamede, N. J. (2002). Classificador de Poemas. *CCTE conference*. De la Rosa, J. et al. (2020). Rantanplan, Fast and Accurate Syllabification and Scansion of Spanish Poetry. *Procesamiento del Lenguaje Natural*. Gamallo, P. et al. (2024). Open Generative Large Language Models for Galician. *Procesamiento del Lenguaje Natural*. Garcia, M. (2021). Exploring the Representation of Word Meanings in Context: A Case Study on Homonymy and Synonymy. *Proc. ACL*. Marco Remón, G., & Gonzalo, J. (2021). Escansión automática de poesía española sin silabación. *Procesamiento del Lenguaje Natural*. Mittmann, A. (2016). *Escansão automática de versos em português* [Tese de doutoramento, Universidade Federal de Santa Catarina]. Mittmann, A. et al. (2019). Aoidos: Una herramienta para análisis automático de versos. *Ene*. Navarro-Colorado, B. (2018). A metrical scansion system for fixed-metre Spanish poetry. *Digital Scholarship in the Humanities*. Pena, X. R (2014). *Historia da literatura galega II. De 1853 a 1916. O Rexurdimento*. Xerais, Pérez Pozo, Á. et al. (2022). A bridge too far for artificial intelligence?: Automatic classification of stanzas in Spanish poetry. *JASIST*. Ruiz Fabo, P. et al. (2017) Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets. *LaTeCH-CLfL*. Vilavedra, D. (1999). *Historia da Literatura Galega*. Galaxia.