

## ANJA, ¿dónde están los encabalgamientos?

**Clara Martínez-Canton**

cimartinez@flog.uned.es  
LINHD, UNED, Spain

**Pablo Ruiz-Fabo**

pablo.ruiz@linhd.uned.es  
LINHD, UNED, Spains

**Elena González-Blanco**

egonzalezblanco@flog.uned.es  
LINHD, UNED, Spain

### Introducción

Encabalgamiento es el desajuste entre la pausa métrica y la sintáctica (Domínguez Caparrós, 2000: 103) que ocurre cuando una unidad de sentido se rompe entre dos versos. Este fenómeno, desde siempre utilizado con distintos fines expresivos (énfasis, ambigüedad, etc.) es difícil de delimitar formalmente.

El estudio más sistemático realizado para su caracterización en español sigue siendo el realizado en su tesis por Quilis (1964). El estudioso experimentó con lecturas de prosa, buscando demostrar qué unidades sintácticas no permiten pausa de sentido en su interior. Basándose en los resultados definió una serie de categorías gramaticales y sintácticas cuya separación en versos distintos produce encabalgamiento. La tipología allí establecida se considera ya clásica. El estudio de Quilis proporciona una definición formal y empírica del fenómeno. Con base en sus reglas se ha creado una herramienta capaz de detectar el encabalgamiento y sus tipos.

Este póster presenta la interfaz ANJA para el análisis automático del encabalgamiento desde una sencilla aplicación web: <http://prf1.org/anja/index/>, desarrollada dentro del proyecto ERC POSTDATA GA- 679528<sup>1</sup>.

### Estado del arte

La naturaleza formal del análisis métrico lo hace un campo propicio para su tratamiento computacional (Birnbaum and Thorsen, 2015; Delente and Renault, 2015). El procesamiento del lenguaje natural (PLN) ofrece muchas posibilidades para la métrica, pues las reglas de definición lingüística permiten llevar a cabo análisis y extracción automática de grandes cantidades de información de corpus textuales.

Para la automatización del análisis métrico en español destacamos los estudios de escansión silábica y acentual de Navarro-Colorado (2017), Agirrezabal (2017) y Gervás (2000). También los trabajos de generación automática de poesía con patrones métricos (Gervás, 2000b) y (Gervás, 2015).

En el campo de las interfaces cabe distinguir entre aquellas que exploran datos de textos ya analizados, recogidos en una base de datos, y aquellas que permiten la entrada y análisis de cualquier poema. Del primer tipo destacamos For Better For Verse<sup>2</sup> (Tucker, 2011) y Database of Czech Metre<sup>3</sup> (Plecháč and Kolár, 2015). Entre las que permiten introducir textos destacamos, en español, la ligada a la herramienta de Navarro-Colorado<sup>4</sup>, que analiza versos endecasílabos. Otros sitios con interfaz de entrada para análisis métrico son Separarensilabas<sup>5</sup> o Lexiquetos<sup>6</sup>. En otras lenguas destacamos Metricalizer<sup>7</sup> (Bobenhausen and Hammerich, 2015) para alemán, Aoidos<sup>8</sup> (Mittmann, 2016) para portugués y español, y RhymeDesign<sup>9</sup> (McCurdy et al., 2015) especializado en rima en inglés.

Una interfaz para el análisis del encabalgamiento representa, sin embargo, una novedad en el campo.

### Herramienta y resultados

El programa de detección del encabalgamiento en español, basado en PLN, se desarrolló en 2016-2017 y fue evaluado sobre dos corpus de test de distintos periodos (Ruiz et al., 2017). ANJA proporciona una interfaz web simple para este programa. El sistema consta de tres componentes: módulo de preprocesado para uniformar el formato de los poemas, pipeline de PLN (basada en IXA Pipes (Agerri et al., 2014) para POS-tagging, constituyentes y dependencias sintácticas) y módulo de detección de encabalgamiento (basado en reglas y diccionarios) y ampliamente documentado en el sitio web<sup>10</sup>. Se ha utilizado esta herramienta para etiquetar un corpus de más de 4000 sonetos alojado y documentado en <https://github.com/postdataproject/disco>.

El código de la herramienta de detección de encabalgamientos está disponible en [https://bitbucket.org/pruizf/anja\\_public/](https://bitbucket.org/pruizf/anja_public/).

### Interfaz gráfica de usuario

ANJA es una interfaz pública y gratuita, alojada en: <http://prf1.org/anja/index/>. Permite cargar los poemas que el

<sup>1</sup> Este trabajo se enmarca dentro del proyecto de investigación Starting Grant Poetry Standardization and Linked Open Data: POSTDATA (ERC-2015-STG-679528), financiado por el European Research Council (ERC) bajo el programa: European Union's Horizon 2020 research and innovation programme, dirigido como Investigador Principal por la profesora Elena González-Blanco, LINHD UNED (<http://postdata.linhd.es/>).

<sup>2</sup> <http://prosody.lib.virginia.edu/>

<sup>3</sup> [http://versologie.cz/v2/web\\_content/](http://versologie.cz/v2/web_content/)

<sup>4</sup> <http://adso.gplsi.es/index.php/es/demostracion/>

<sup>5</sup> <http://www.separarensilabas.com/index.php>

<sup>6</sup> <http://lexiquetos.org/silio/>

<sup>7</sup> <https://metricalizer.de/en/metrikanalyse/poem>

<sup>8</sup> <http://aoidos.ufsc.br/>

<sup>9</sup> <http://www.sci.utah.edu/~nmccurdy/rhymeDesign/>

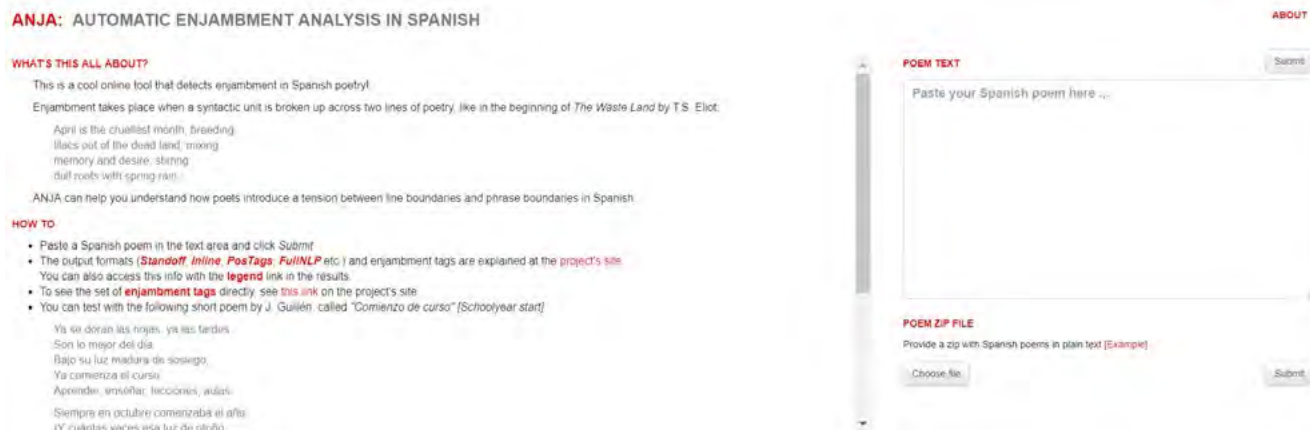
<sup>10</sup> <https://sites.google.com/site/spanishenjambment/>

usuario decida y analizarlos en el momento. También ofrece la carga de archivos ZIP que contengan archivos en texto plano.

La interfaz de usuario está construida con el framework Django (Python), con las plantillas de Bootstrap 3. Las vistas de Django se llaman con AJAX para poblar los elementos de la UI. Para el análisis de PLN, Django

accede a servicios web Java (IXA Pipes) implementados en nuestro servidor.

ANJA presenta dos ventanas de navegación (Fig. 1), la principal, para introducir poemas, a la derecha y, a la izquierda, una mínima guía de uso que explica su funcionamiento y enlaza a la web del proyecto.



Captura de ANJA

Los resultados se ofrecen dos formatos: *Standoff* (tipo de encabalgamiento y línea), e *Inline* (etiquetado gramatical y tipo de encabalgamiento por línea, ver Fig.

2 para *Inline*). Las anotaciones PLN en que se basa en sistema se ofrecen en las pestañas *PosTags* (etiquetas gramaticales) y *FullNLP* (pipeline completa).

El enlace **legend**<sup>11</sup> da acceso a la leyenda que explica los tipos de encabalgamiento, las etiquetas gramaticales y otras convenciones de representación:

| # | Text  | Position | Enjambment Type |
|---|---|----------|-----------------|
| 1 | {Ya A} {se Q} {doran V} {las D} {hojas N} {, O} {ya A} {las D} {tardes N} | B        | ex_subj_verb    |
| 2 | {Son V} {lo D} {mejor G} {del P} {día N}                                  | I        | ex_subj_verb    |
| 3 | {Bajo P} {su D} {luz N} {madura G} {de P} {sosiego N} {, O}               | O        |                 |
| 4 | {Ya A} {comienza V} {el D} {curso N} {, O}                                | O        |                 |
| 5 | {Aprender V} {, O} {enseñar V} {, O} {lecciones N} {, O} {aulas N} {, O}  | O        |                 |
| 6 | {Siempre A} {en P} {octubre O} {comenzaba V} {el D} {año N} {, O}         | O        |                 |
| 7 | {i O} {Y C} {cuántas Q} {veces N} {esa D} {luz N} {de P} {otoño N}        | O        |                 |

Anotaciones de encabalgamiento en formato *Inline*

La existencia de una aplicación web simple para la utilización esta herramienta la hace accesible para una gama mucho más amplia de usuarios.

## References

Agerri, R., Bermudez, J. and Rigau, G. (2014). IXA pipeline: Efficient and Ready to Use Multilingual NLP tools. *Proceedings of LREC 2014, the 9th International Language Resources and Evaluation Conference*, vol. 2014. Reykjavik, Iceland, pp. 3823–3828

<sup>11</sup> <https://sites.google.com/site/spanishenjambment/legend>

- [http://www.lrec-conf.org/proceedings/lrec2014/pdf/775\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/775_Paper.pdf) (accessed 20 April 2017).
- Agirrezabal, M. (2017). Automatic Scansion of Poetry San Sebastián/Donosti: Universidad del País Vasco.
- Birnbaum, D. J. and Thorsen, E. (2015). Markup and meter: Using XML tools to teach a computer to think about versification. *Balisage: The Markup Conference* <http://www.balisage.net/Proceedings/vol15/print/Birnbaum01/BalisageVol15-Birnbaum01.html> (accessed 22 April 2017).
- Bobenhausen, K. and Hammerich, B. (2015). Métrique littéraire, métrique linguistique et métrique algorithmique de l'allemand mises en jeu dans le programme Metricalizer2. *Langages*(3): 67–88.
- Delente, É. and Renault, R. (2015). Outils et métrique: un tour d'horizon. *Langages*(3): 5–22.
- Domínguez Caparrós, J. (2000). *Métrica Española*. Madrid: Síntesis.
- Gervás, P. (2000a). A Logic Programming Application for the Analysis of Spanish Verse. *Computational Logic—CL 2000*. Berlin: Springer Berlin Heidelberg, pp. 1330–44.
- Gervás, P. (2000b). Wasp: Evaluation of different strategies for the automatic generation of spanish verse. *Proceedings of the AISB-00 Symposium on Creative & Cultural Aspects of AI*. pp. 93–100 [https://www.researchgate.net/profile/Pablo\\_Gervas/publication/228609235\\_Wasp\\_Evaluation\\_of\\_different\\_strategies\\_for\\_the\\_automatic\\_generation\\_of\\_spanish\\_verse/links/00b4952aada6407047000000.pdf](https://www.researchgate.net/profile/Pablo_Gervas/publication/228609235_Wasp_Evaluation_of_different_strategies_for_the_automatic_generation_of_spanish_verse/links/00b4952aada6407047000000.pdf) (accessed 22 April 2017).
- Gervás, P. (2015). Tightening the Constraints on Form and Content for an Existing Computer Poet. *AISB Convention 2015* <http://eprints.sim.ucm.es/37000/> (accessed 22 April 2017).
- McCurdy, N., Srikumar, V. and Meyer, M. (2015). Rhyme-design: A tool for analyzing sonic devices in poetry. *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*. pp. 12–22.
- Mittmann, A. (2016). Escansão automática de versos em português. <https://repositorio.ufsc.br/handle/123456789/175819>.
- Navarro-Colorado, B. (2017). A metrical scansion system for fixed-metre Spanish poetry. *Digital Scholarship in the Humanities* doi:10.1093/llc/fqx009. <https://academic.oup.com/dsh/article-abstract/doi/10.1093/llc/fqx009/3064339/A-metrical-scansion-system-for-fixed-metre-Spanish> (accessed 19 April 2017).
- Plecháč, P. and Kolár, R. (2015). The Corpus of Czech Verse. *Studia Metrica et Poetica*, 2(1): 107–118.
- Quilis, A. (1964). *Estructura Del Encabalgamiento En La Métrica Española*. Consejo Superior de Investigaciones Científicas, patronato' Menéndez y Pelayo,' Instituto' Miguel de Cervantes,'
- Ruiz Fabo, P., Bermúdez Sabel, H., Martínez Cantón, C. I., González-Blanco, E. and Navarro-Colorado, B. (2018). The Diachronic Spanish Sonnet Corpus (DISCO): TEI and Linked Open Data Encoding, Data Distribution and Metrical Findings. *Humanidades Digitales 2018 (DH 2018)*. Ciudad de México, México.
- Ruiz Fabo, P., Bermúdez-Sabel, H., Martínez Cantón, C. I. and Calvo Tello, J. (2017). *Diachronic Spanish Sonnet Corpus (DISCO)*. Madrid: UNED. Madrid <https://doi.org/10.5281/zenodo.1012567>.
- Ruiz, P., Martínez Cantón, C., Poibeau, T. and González-Blanco, E. (2017). Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets. *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Association for Computational Linguistics, pp. 27–32.
- Tucker, H. F. (2011). Poetic data and the news from poems: A for better for verse memoir. *Victorian Poetry*, 49(2): 267–281.

---

## Combining String Matching and Cost Minimization Algorithms for Automatically Geocoding Tabular Itineraries

**Rui Santos**

[rui@rui.santos.pt](mailto:rui@rui.santos.pt)  
IST and INESC-ID,  
University of Lisbon, Portugal

**Bruno Emanuel Martins**

[bruno.g.martins@ist.utl.pt](mailto:bruno.g.martins@ist.utl.pt)  
IST and INESC-ID,  
University of Lisbon, Portugal

**Patricia Murrieta-Flores**

[p.murrietaflores@chester.ac.uk](mailto:p.murrietaflores@chester.ac.uk)  
Digital Humanities Research Center,  
University of Chester, United Kingdom

Historical itineraries, often accessible as tables or as sequential lists of names for the places visited in the context of a particular journey, are abundant resources and also important objects of study for Humanities scholars, providing 'snapshots' of particular socio-cultural events, insights into the development of human mobility, and invaluable information related to the establishment of road networks. Well-known examples include the 3rd century *Itinerarium Antonini Augusti* or the *Itinerarium Burdigalense*, written between the 8th and 10th centuries, among others. Many historical manuscripts and/or transcriptions containing information on itineraries, dating from the medieval period to the 20th century, are nowadays available in digital formats, through initiatives such as Europeana or the Internet Archive, or in the context of Digital Humanities projects like Pelagios.

Few historical tabular itineraries are nonetheless directly associated with map-based representations and, in many cases, there is little information on the actual routes