

Cuestiones y problemas en la anotación métrica de un corpus de poesía castellana

M. A. Herrero, **B. Navarro-Colorado**, M. Ribes Lafoz, A.
L. Prieto de Paula, F. Chico Rico. Universidad de Alicante.

Humanidades Digitales Hispánicas - Málaga, 20/X/2017

Proyecto ADSO

Objetivos

Análisis “distante” del sonetos castellano del Siglo de Oro.

- Rasgos recurrentes de la métrica y la semántica del soneto áureo.

Objetivos

Análisis “distante” del sonetos castellano del Siglo de Oro.

- Rasgos recurrentes de la métrica y la semántica del soneto áureo.

Recursos:

- Un amplio corpus de sonetos.
- Un sistema automático de escansión.

Corpus de sonetos

45 autores, 5078 sonetos, más de 71.000 versos.

- Fuente: Biblioteca del Soneto de la BVMC.
- **Descarga** (licencia *creative commons*):
<https://github.com/bncolorado/CorpusSonetosSigloDeOro>
- Consulta: <http://goldenage.cervantesvirtual.com>

Corpus de sonetos

XML-TEI

Anotación métrica: secuencia de sílabas átonas y tónicas.

```
<l n="1" met="--+----+--">
```

 Cuando me paro a contemplar mi estado

```
</l>
```

Sistema de escansión

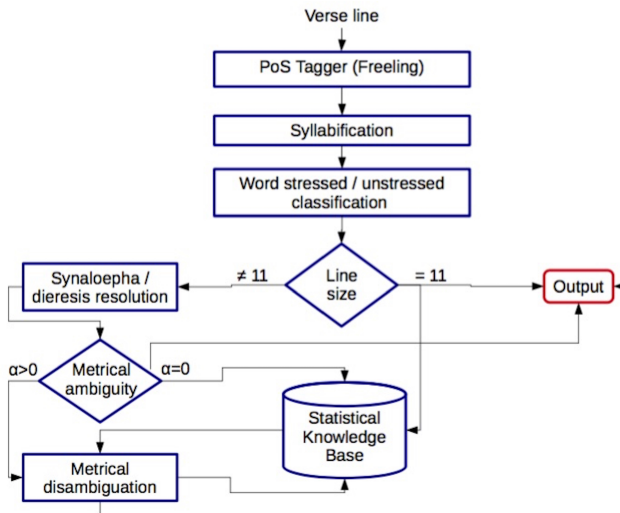


Figure 1: Sistema de escansión

Sistema de escansión

- Descarga (licencia GNU):
<https://github.com/bncolorado/adsoScansionSystem>
- Demostración:
<http://adso.gplsi.es/index.php/es/demostracion/>

Proceso de anotación

- ① Análisis automático de todo el corpus
- ② Extracción de versos potencialmente erróneos (920 versos):
 - Cantidad errónea (!= 11).
 - Métrica poco frecuente (< 10).
 - Detección errores sistemáticos ocultos (ver luego).
- ③ Corrección y validación manual
 - 826 versos (89,7%)

Análisis de errores

Tipos de errores

- Textuales
- Métricos
- Interpretativos (ambigüedad)

Errores textuales

① Erratas:

Ejemplo

F. de Herrera (255) "¡Ay dura, ay importuna, ay triste ausencia!"

```
<l n="10" met="-+-+--+--+--">  
    pues un espíritu en dos mitades parte.</l>
```

```
<l n="10" met="-+-+--+--+--">  
    pues un espirtu en dos mitades parte.</l>
```

Errores textuales

② Errores textuales en origen

Ejemplo

Mira de Amescúa (18) "Un mar y una garita me hacen roncha;"

```
<l n="8" met="+++--+-+--+-">
```

```
no des troncha, si des troncha, no destronca.
```

```
</l>
```

Errores textuales

③ Variantes textuales

Ejemplo

Juan de Salinas (11) "¡Oh carga de un Prelado religioso!"

```
<l n="6" met="---+---+++-">  
    si disimula, de blando es muy sobrado;</l>
```

(Moíño Sánchez, 2007):

```
<l n="6" met="-+---+---+-">  
    si en algo disimula, acobardado</l>
```

Errores textuales

Se mantiene la edición de la BVMC:

- Declaración en *TEI header*.

Se mantienen los errores métricos derivados de errores textuales.

- Sólo se corrigen erratas textuales muy evidentes.

Errores análisis métrico

Errores del sistema automático de escansión:

- Versos no endecasílabos (en estrambote).
- Error en el análisis categorial de las palabras.
- Error en la resolución de diéresis o sinalefas.

Errores análisis categorial

Ejemplos

Fernando de Herrera (20) "Al puro ardor que vibran mis estrellas,"

```
<l n="2" met="(+)++--++--++--">
```

```
do Amor sus rayos templa en dulce fuego</l>
```

Errores análisis categorial

Ejemplos

Fernando de Herrera (20) "Al puro ardor que vibran mis estrellas,"

```
<l n="2" met="(+) +--+--+--+--">
  do Amor sus rayos templa en dulce fuego</l>
```

Francisco de Quevedo (212) "Díceme, don Jerónimo, que dices"

```
<l n="1" met="+--(+)-+----+--">
  Díceme, don Jerónimo, que dices</l>
```

Errores en la resolución de sinalefas y diéresis

Ejemplo

Lope de Vega (577) "Pasaba el claro Eveno a Deianira"

```
<l n="9" met="++---+---+--">
```

```
No| fue|ra| tan| cruel| mi| ai|ra|da| suer|te
```

```
</l>
```

```
<l n="9" met="++---+---+--">
```

```
No| fue|ra| tan| crü|el| mi_ai|ra|da| suer|te
```

```
</l>
```

Versos con ambigüedad métrica

Versos con ambigüedad métrica

Vocativos:

Lope de Vega (260) "Señora mía, si de vos ausente"

```
<l n="1" met="-(+)-+----+--">
```

```
Señora mía, si de vos ausente</l>
```

Gutierre de Cetina (178) "Triste avecilla que te vas quejando"

```
<l n="1" met="(+)---+----+--">
```

```
Triste avecilla que te vas quejando</l>
```

Versos con ambigüedad métrica

“Mientras” (adverbio vs. conjunción):

Francisco de la Torre (57) "Santa madre de amor, que el yerto suelo"

```
<l n="5" met="(-)----+--+--">
  mientras al descubierto y raso cielo</l>
<l n="6" met="+--+--+---+--">
  pacen sus vacas hierba floreciente,</l>
```

Ampliación y nuevos desarrollos

Nuevo modelo de representación enriquecido

Ahora incluye:

- Enlace a la url de donde se ha obtenido el texto (en el *TEI Header*)
- Tipo de poema.
- Separación silábica del verso.
- Lema y categoría gramatical de cada palabra.

Nuevo modelo de representación enriquecido

Calderón de la Barca “Aunque la persecución”

```
<lg cert="high" type="Décima">
```

```
...
```

```
<l met="-|-|3|-|5|-|7|-|" n="2">
```

```
  <span type="raw">de la envidia tema el sabio,</span>
```

```
  <w lemma="de" type="SP">de|</w>
```

```
  <w lemma="el" type="DAOFSO">la</w>
```

```
  <w lemma="envidia" type="NCFS000">en|vi|dia|</w>
```

```
  <w lemma="tema" type="NCMS000">te|ma</w>
```

```
  <w lemma="el" type="DAOMSO">el|</w>
```

```
  <w lemma="sabio" type="NCMS000">sa|bio|</w>
```

```
</l>
```

Ampliación del corpus

Aprox. 700 poemas nuevos (+100.000 versos):

- Romances
- Octavas reales
- Églogas
- Liras
- Canciones
- Décimas
- Coplas
- Villancicos etc.

Ampliación sistema escansión

Adaptación del sistema de escansión a cualquier tipo de verso.

- Técnicas de aprendizaje automático:
- La cantidad silábica depende de los versos anteriores (secuencia)
- Modelos Ocultos de Markov (HMM)

Conclusiones

Conclusiones

Proyecto ADSO: cuestiones y problemas en la anotación del corpus de sonetos (y ampliación).

- Modelo de representación métrica y sistemas de escansión.
- Problemas:
 - Textuales
 - Métricos: categorías gramaticales, sinalefas...
 - Ambigüedad.
- Situación actual:
 - Enriquecimiento del modelo de representación.
 - Ampliación del corpus con otros tipos de corpus
 - Ampliación del sistema de escansión

GRACIAS

Bibliografía

- JOCKERS, Matthew L. (2013), *Macroanalysis. Digital Methods and Literary History*, University of Illinois Press, Illinois (EEUU).
- MOÍÑO SÁNCHEZ, PABLO (2007) «¿Redondilla o quadra? Notas sobre unos versos del siglo XVII» *Lectura y signo* (2), pp. 89-98.
- MORETTI, Franco (2007), *La literatura vista desde lejos*, Marbot ediciones, Barcelona.
- NAVARRO COLORADO, Borja (2015), «A computational linguistic approach to Spanish Golden Age Sonnets: metrical and semantic aspects», *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver (EEUU), junio 2015.

Bibliografía

- NAVARRO COLORADO, Borja (2017), «A Metrical Scansion System for Fixed-Metre Spanish Poetry», *Digital Scholarship in the Humanities*. DOI: <https://doi.org/10.1093/llc/fqx009>
- PADRÓ, Luis y STANILOVSKY, E. (2012), «FreeLing 3.0: Towards Wider Multilinguality», *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Estambul, Mayo 2012.
- QUILIS, Antonio (1984) *Métrica española*, Ariel, Barcelona.
- TEI CONSORTIUM (ed) (2016), *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Version 3.1.0. Last modified 15th December 2016. TEI Consortium. EN <http://www.tei-c.org/Guidelines/P5/> (Fecha de consulta: 19 de abril de 2017).

Agradecimientos

Trabajo financiado por la **Fundación BBVA**, ayudas a equipos de investigación científica 2016, proyecto “Análisis distante de base computacional del soneto castellano del Siglo de Oro”.

<http://adso.gplsi.es/index.php/es/proyecto-adso/>