Poetry and Digital Humanities making interoperability possible in a divided world of digital poetry: POSTDATA project

Salvador Ros
UNED-LINHD
Elena González-Blanco García
COVERWALLET – LINHD

Galway, 9 December 2018
The “digital gap”
POSTDATA Vision is to ease humanists research through its digital transformation

<table>
<thead>
<tr>
<th>Traditional Research</th>
<th>Digital Humanities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counting syllables with fingers</td>
<td>Automatic syllable counter</td>
</tr>
<tr>
<td>Comparing 2-3 poems at the same time</td>
<td>Comparing thousands of poems in parallel</td>
</tr>
<tr>
<td>Looking for similar evidences by reading all papers and books</td>
<td>Finding multiple evidences through one search engine</td>
</tr>
<tr>
<td>Individual research</td>
<td>Collaborative, interdisciplinary research</td>
</tr>
</tbody>
</table>
The three POSTDATA Axes

- **Semantic Infrastructure**
  - Linked open data

- **Virtual Research Environment**
  - Digital Scholarly Editing

- **Poetry Lab**
  - Natural language processing
There is not an ontological model for poetry
It is necessary to standardize specific fields (rhyme, metrical scheme)
It is necessary to map controlled vocabularies across different languages
But, what does define a POEM?

Author
Title
Incipit
Manuscript
Post quem
Ante quem
Language
Topics
Edition
Online edition
Work

Isometrism
Isostrophism
Metrical
scheme
Rhyme scheme
Rhyme
Musical
notation
Number of
Stanzas
Number of
lines
Poetic form

Estuans intrinsecus
in amaritudine
factus de materia
folio sum similis,

Cum sit enim proprium
supra petram ponere
stultus ego comparor
sub eodem aere

Feror ego veluti
ut per vias aeris
non me tenent vincula,
quero mei similes

Carmina Burana, 191
DOMAIN MODEL FOR EUROPEAN POETRY

RPHA - Hungary
MedDB2 – Galicia
ReMetCa - Spain

BedT - Italy
Naetebus - France
Skaldic - Scandinavia

Cantigas Oxford
Analecta Hymnica Latin
Mirabile - Italy

The puzzle of digital poetic repertoires
POSTDATA looks for an abstract model with common elements across the different poetic systems.
DOMAIN MODEL FOR EUROPEAN POETRY

key concepts:
  • standardisation and
  • interoperability

The most advantageous paradigm
  • Linked Open Data
Poetry Lab : NLP tools

• Tools for (automatic) poetry analysis
• Largely based on Natural Language Processing
• Insofar as literary phenomena in domain model rely on linguistic traits, NLP can help automate their annotation
Content analysis:
- Medieval Spanish Named Entity Recognition (Díez Platas et al., 2017)

Phonetics and metrics:
- Lexical Syllabification (prereq for scansion)

Other stylistic traits:
- Enjambment detection (Ruiz, Martínez et al., 2017)
Poetry Lab: Syllabification

- A prerequisite for scansion
- Syllable boundary detection + pos-tagging to determine stressability (weak vs. strong categories)

<table>
<thead>
<tr>
<th>#</th>
<th>Text</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ya se doran las hojas, ya las tardes</td>
<td>Ya</td>
<td>se</td>
<td>do</td>
<td>ran</td>
<td>las</td>
<td>ho</td>
<td>jas</td>
<td>ya</td>
<td>las</td>
<td>tar</td>
<td>des</td>
</tr>
<tr>
<td>2</td>
<td>Son lo mejor del día</td>
<td>Son</td>
<td>lo</td>
<td>me</td>
<td>jor</td>
<td>del</td>
<td>di</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Bajo su luz madura de sosiego,</td>
<td>Ba</td>
<td>jo</td>
<td>su</td>
<td>luz</td>
<td>ma</td>
<td>du</td>
<td>ra</td>
<td>de</td>
<td>so</td>
<td>sie</td>
<td>go</td>
</tr>
</tbody>
</table>
Poetry Lab: Enjambment

• Mismatch between metrical and syntactic units
• Defined in syntactic terms
• Can be detected based on pos-tagging / parsing
• Definition/detection not trivial
• Previously unanalyzed with automatic means
Poetry Lab: Enjambment

- Enjambment takes place when particularly cohesive syntactic units are broken up by a metrical pause (across lines or hemistichs)

- Example
  1. Prefiero
  2. quedarme desnuda //
  3. de amigo insincero

  1. I’d rather
  2. be naked //
  3. of fake friend

INTERPRETATIONS

- a) naked (no clothes)
- b) naked of fake friend (without such a person)
Poetry Lab: NERC

• Hispanic Medieval Tagger (Díez Platas et al., 2017)
• Medieval Spanish NERC
• Person, location, organization, role names
• Normalization of historical spellings
• Geolocalization
• TEI output
Poetry Lab: NERC
sros@scc.uned.es
http://postdata.linhd.es