



European
Research
Council



Rantanplan

Fast and Accurate Syllabification and Scansion of Spanish Poetry

Javier de la Rosa (versae@linhd.uned.es)

Álvaro Pérez (alvaro.perez@linhd.uned.es)

Laura Hernández (laura.hernandez@scc.uned.es)

Salvador Ros (sros@scc.uned.es)

Elena González Blanco (egonzalezblanco@faculty.ie.edu)

Laboratorio de Innovación en Humanidades Digitales (LINHD), UNED, Spain

Postdoctoral Researcher at ERC POSTDATA Project (Poetry Standardization and Linked Open Data)



Outline



POSTDATA

Poetry Standardization
and Linked Open Data

- Introduction
- Fast Scansion
 - PoS
 - Syllabification
 - Stress and phonological groups
 - Metrical adjustment
- Evaluation
 - Syllabification
 - Scansion
- Limitations
- Conclusions

- Syllabification
 - *tambor*
 - *planta*
 - *plátano*

- Syllabification
 - *tam-bor*
 - *plan-ta*
 - *plá-ta-no*

Introduction



POSTDATA

Poetry Standardization
and Linked Open Data

- Stress
 - *tam-**bor***
 - ***plan**-ta*
 - ***plá**-ta-no*

- Stress
 - *tam-**bor***
 - *oxytone* (aguda)
 - ***plan**-ta*
 - *paroxytone* (llana)
 - ***plá**-ta-no*
 - *proparoxytone* (esdrújula)

- Scansion
 - *tam-**bor***
 - *oxytone* (aguda)
 - ***plan**-ta*
 - *paroxytone* (llana)
 - ***plá**-ta-no*
 - *proparoxytone* (esdrújula)

- Scansion
 - *se-gún*
 - *oxytone* (aguda)
 - *an-te*
 - *paroxytone* (llana)

- Scansion
 - *se-gún*
 - ~~*oxytone*~~ (aguda)
 - *an-te*
 - ~~*paroxytone*~~ (llana)

- Scansion
 - *tam-**bor***
-+
 - ***plan**-ta*
+-
 - ***plá**-ta-no*
+--

- Scansion

- *tam-**bor***

-+ → -+-

- ***plan**-ta*

+ - → + -

- ***plá**-ta-no*

+ -- → + -

- Synalepha

Cuando el alba me despierta

*Cuan-do el-al-ba me des-**pier**-ta*

— — + — — — + — 8

(Miguel de Unamuno)

- Syneresis

y al ver sonreír los astros, me prosterno

*y al ver son-**re-ír** los as-tros, me pros-ter-no*

— — — + — + — — — + — 11

(Manuel de Montoliu)

- Dieresis

en cánticos y nácares süaves

*en cán-ti-cos y ná-ca-res **sü-a**-ves*

– + – – – + – – – + – 11

(Fray Jerónimo de San José)

Fast scansion



POSTDATA

Poetry Standardization
and Linked Open Data

- 1) PoS tagging
- 2) Syllabification
- 3) Stress and phonological groups
- 4) Metrical adjustment

Fast scansion



POSTDATA

Poetry Standardization
and Linked Open Data

- PoS tagging
- Stress and phonological groups
 - spaCy model (es_core_news)
 - AnCora
 - Freeling
 - spaCy-affixes

- Syllabification

Ríos Mestre (1998), Caparrós (1993) and Navarro Tomás (1991)

- 1) Pre-syllabification
- 2) Letter groups (regex)
- 3) Post-syllabification

- Stress and phonological groups
 - Synalephas and syneresis are applied (more common than dieresis)
 - Stress is extracted and transferred

me ama

*me-**a**-ma*

***mea**-ma*

- Pattern is returned

- Metrical adjustment
 - If metrical length is known
AND
length returned is lower than expected
 - Then undo combinations of
 - First, synalephas
 - Secondly, syneresis
 - Finally, both

Fast scansion



POSTDATA

Poetry Standardization
and Linked Open Data

Algorithm 1: Scansion procedure

Input: A sequence \mathcal{W} of words
 $\langle w_1, w_2, \dots, w_n \rangle$

Input: A value $length$ for the
metrical length expected
(optional)

Output: A sequence $\langle s_1, s_2, \dots, s_{\mathcal{L}} \rangle$
of booleans expressing the
metrical pattern

```
1 for  $w_i \in \mathcal{W}$  do
2   |  $tag_i \leftarrow \text{pos}(w_i)$ 
3   |  $syllables_i \leftarrow \text{syllabify}(w_i)$ 
4   |  $stresses_i \leftarrow \text{stress}(syllables_i, tag_i)$ 
5 end
6  $groups \leftarrow \text{phonological}(syllables,$ 
   |  $stresses)$ 
7  $pattern \leftarrow \text{transform}(groups)$ 
8 if  $length$  then
9   | while  $|pattern| < length$  do
10  | |  $g \leftarrow \text{generate\_phonological}(\mathcal{W})$ 
11  | |  $pattern \leftarrow \text{transform}(g)$ 
12  | end
13 end
14 return  $pattern$ 
```

Evaluation: Syllabification

- EDFU (100k words)

Method	Accuracy (%)
Navarro-Colorado	98.35
Agirrezabal	98.06
Rantanplan (ours)	99.99

Evaluation: Scansion

- 1400 hendecasyllables

Method	Accuracy (%)	Time
Gervás	70.88	N/A
Navarro-Colorado	94.45	2,356 seconds
Agirrezabal	90.84	N/A
Rantanplan (ours)	96.23	21 seconds

Evaluation: Scansion

- 10k hendecasyllables

Method	Accuracy (%)	Time
Gervás	67.56	N/A
Navarro-Colorado	90.89	16,787 seconds
Rantanplan (ours)	92.75	53 seconds

Evaluation: Scansion

- 4300 mixed-metre verses

Method	Accuracy (%)	Time
Navarro-Colorado	49.38	7,484 seconds
Rantanplan (ours)	65.02	27 seconds

Limitations



POSTDATA

Poetry Standardization
and Linked Open Data

- Heuristic based
- Not contemporary corpora
- Statistical PoS tagger

Conclusions



POSTDATA

Poetry Standardization
and Linked Open Data

- New syllabification corpus
- Apply synalephas and syneresis
- State of the art results

- Usage example



```
from rantanplan import get_scansion

poem = """Me gustas cuando callas porque estás como ausente,
y me oyes desde lejos, y mi voz no te toca.
Parece que los ojos se te hubieran volado
y parece que un beso te cerrara la boca.

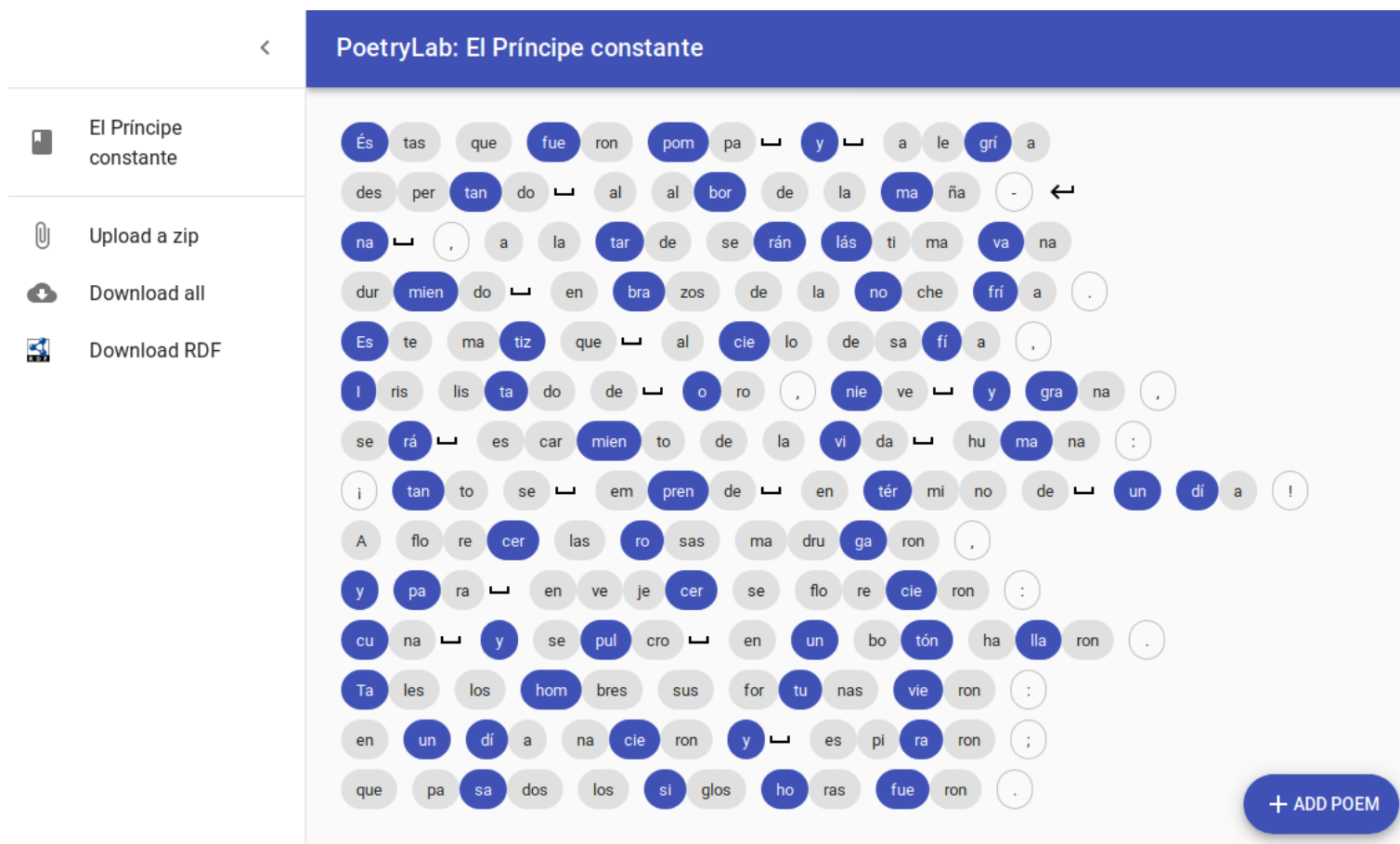
Como todas las cosas están llenas de mi alma
emerges de las cosas, llena del alma mía.
Mariposa de sueño, te pareces a mi alma,
y te pareces a la palabra melancolía."""

get_scansion(poem)
```

- Usage example

```
    [{{'tokens': [{{'word': [{{'syllable': 'Me',  
      'is_stressed': False,  
      'is_word_end': True}],  
      'stress_position': 0}},  
    {{'word': [{{'syllable': 'gus', 'is_stressed': True},  
      {{'syllable': 'tas', 'is_stressed': False, 'is_word_end': True}],  
      'stress_position': -2}},  
    {{'word': [{{'syllable': 'cuan', 'is_stressed': False},  
      {{'syllable': 'do', 'is_stressed': False, 'is_word_end': True}],  
      'stress_position': 0}},  
    ...]},  
  'phonological_groups': [{{'syllable': 'Me',  
    'is_stressed': False,  
    'is_word_end': True},  
    {{'syllable': 'gus', 'is_stressed': True},  
    {{'syllable': 'tas', 'is_stressed': False, 'is_word_end': True},  
    {{'syllable': 'cuan', 'is_stressed': False},  
    {{'syllable': 'do', 'is_stressed': False, 'is_word_end': True},  
    ...]},  
  'rhythm': {{'stress': '-+----+----+---+', 'type': 'pattern', 'length': 14}},  
  ...
```

- Usage example with the UI



The screenshot displays the PoetryLab interface for the poem "El Príncipe constante". The poem text is presented in a grid of words, with some words highlighted in blue. The text is as follows:

És tas que fue ron pom pa y a le gri a
des per tan do al al bor de la ma ña -
na , a la tar de se rán lás ti ma va na
dur mien do en bra zos de la no che fri a .
Es te ma tiz que al cie lo de sa fi a ,
I ris lis ta do de o ro , nie ve y gra na ,
se rá es car mien to de la vi da hu ma na :
i tan to se em pren de en tér mi no de un dí a !
A flo re cer las ro sas ma dru ga ron ,
y pa ra en ve je cer se flo re cie ron :
cu na y se pul cro en un bo tón ha lla ron .
Ta les los hom bres sus for tu nas vie ron :
en un dí a na cie ron y es pi ra ron ;
que pa sa dos los si glos ho ras fue ron .

On the left side, there is a sidebar with the following options:

- El Príncipe constante
- Upload a zip
- Download all
- Download RDF

At the bottom right, there is a button labeled "+ ADD POEM".

Availability



POSTDATA

Poetry Standardization
and Linked Open Data

- API:
 - <http://postdata.uned.es:5000/openapi.json>
 - <http://postdata.uned.es:5000/ui>
- Frontend
 - <http://postdata.uned.es/poetrylab/>

Reproducibility

- Repo:
<https://github.com/linhd-postdata/rantanplan-evaluation>
- Try it:
<https://mybinder.org/v2/gh/linhd-postdata/rantanplan-evaluation/master>

Evaluation for Rantanplan v0.4.3

This repository contains Jupyter Notebooks with the evaluation of the scansion system Rantanplan, presentend in the paper "Rantanplan, Fast and Accurate Syllabification and Scansion of Spanish Poetry" Journal the Spanish Natural Language Processing Association (Revista de Procesamiento del Lenguaje Natural), volume 65, 2020.

The evaluation metrics for Rantanplan are split as follows:

- **Fixed metre** contains the evaluation on the ~1400 and ~10000 hendecasyllabic poems.
- **Mixed metre** contains the evaluation on the ~4300 mixed-metre poems.
- **Syllabification** contains the evaluation on the ~100k syllabified words (see [EDFU](#)).

Try it on binder: 

Thanks



POSTDATA

Poetry Standardization
and Linked Open Data

versae@linhd.uned.es
[@versae](#)



POSTDATA

Poetry Standardization
and Linked Open Data