

Automatic Enjambement Detection as a New Source of Evidence in Spanish Versification

Plotting Poetry / Machiner la poésie, Basel 2017

Clara Martínez Cantón,¹ Pablo Ruiz Fabo,¹

Elena González-Blanco,¹ Thierry Poibeau²

¹ Dept. of Spanish and Literary Theory (UNED, Madrid)

² LATTICE Lab (ENS, Paris)



Talk Outline

1. Problem: Defining enjambment
2. Enjambment in Spanish
3. Enjambment detection system
 - a. System description
 - b. Evaluation: Corpus and results
4. Relevance for literary scholarship
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
5. Conclusion and outlook

Talk Outline

- 1. Problem: Defining enjambment**
2. Enjambment in Spanish
3. Enjambment detection system
 - a. System description
 - b. Evaluation: Corpus and results
4. Relevance for literary scholarship
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
5. Conclusion and outlook

Defining enjambment

- Device whereby a syntactic unit is broken up across two lines of poetry.
- An unnatural “cut” ensues, which can result in stylistic effects (emphasis, contrast, double interpretations)

Enjambment example

EXAMPLE

(1) SPANISH

1. Prefiero
2. quedarme **desnuda** //
3. **de amigo insincero**

GLOSS

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

Enjambment example

EXAMPLE

(1) SPANISH

1. Prefiero
2. quedarme **desnuda** //
3. **de amigo insincero**

GLOSS

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*)

Enjambment example

EXAMPLE

(1) SPANISH

1. Prefiero
2. quedarme **desnuda** //
3. **de amigo insincero**

GLOSS

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*)

Enjambment example

EXAMPLE

(1) SPANISH

1. Prefiero
2. quedarme **desnuda** //
3. **de amigo insincero**

GLOSS

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*)

Current debates

- Which syntactic elements can trigger this “unnatural” effect when split?
- Are syntactic criteria the best way to define contexts for these stylistic effects?
- What are the stylistic effects possible?

Intended contribution

- Gathering variety of enjambment examples systematically, thanks to automatic detection
- Shed light on the limitations of current definitions of the phenomenon
- Provide quantitative evidence of the use of enjambment across periods, genres or schools

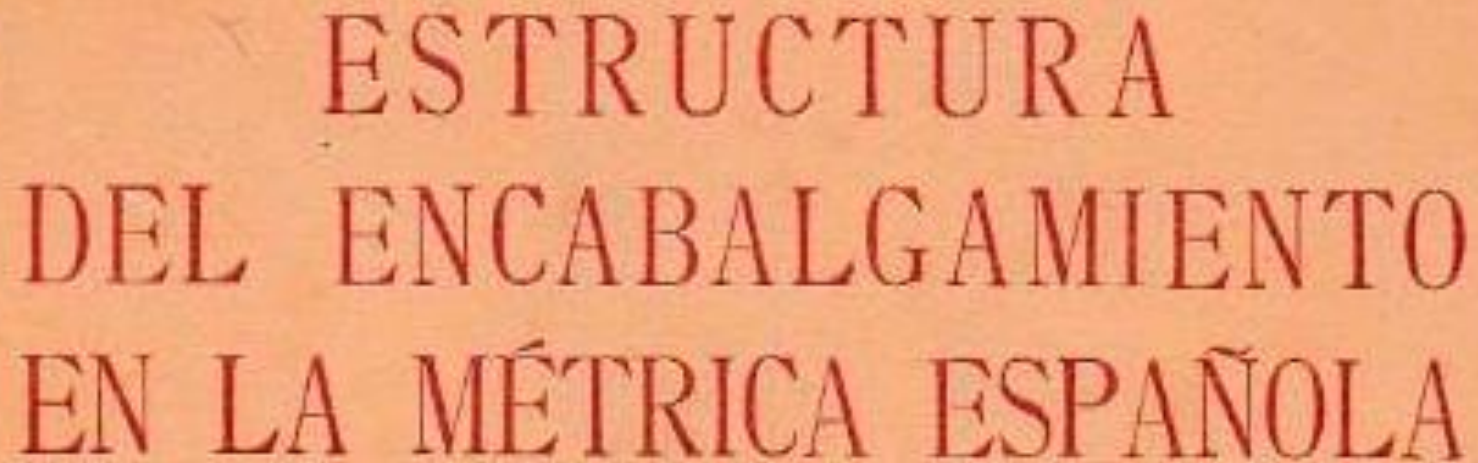
Talk Outline

1. Problem: Defining enjambment
- 2. Enjambment in Spanish**
3. Enjambment detection system
 - a. System description
 - b. Evaluation: Corpus and results
4. Relevance for literary scholarship
 - a. Refining definitions
 - b. Quantitative studies on a large corpus
5. Conclusion and outlook

Enjambment in Spanish



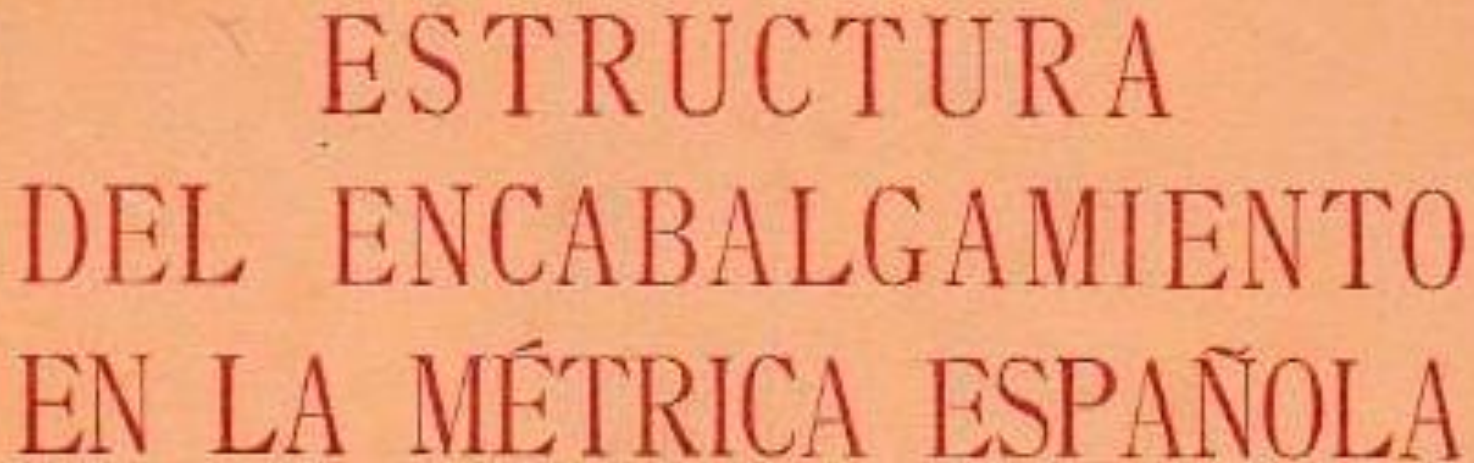
Enjambment in Spanish



ESTRUCTURA
DEL ENCABALGAMIENTO
EN LA MÉTRICA ESPAÑOLA

- Quilis (1964). Estructura del encabalgamiento en la métrica española. Madrid, CSIC.
- Martínez Cantón, C. (2011). Ch. 4 « La Pausa » *Métrica y estética de Antonio Colinas*. Sevilla. Padilla Libros.

Enjambment in Spanish



ESTRUCTURA
DEL ENCABALGAMIENTO
EN LA MÉTRICA ESPAÑOLA

- Quilis (1964). Estructura del **encabalgamiento** en la métrica española. Madrid, CSIC.
- Martínez Cantón, C. (2011). Ch. 4 « La Pausa » *Métrica y estética de Antonio Colinas*. Sevilla. Padilla Libros.

Definition in Quilis (1964)

- Quilis (1964) sees two types of definitions in the literature:
 - Based on the notion of “sense continuation” after a metrical pause
 - Based on the idea of breaking up syntactic components
- He focused on a syntactic definition, finding it more amenable to formalization/testing

Quilis' production experiments

- Enjambment takes place when *particularly cohesive syntactic units* are broken up by a metrical pause (across lines or hemistichs)
- Cohesive units experimentally determined:
 - via **literary prose reading**, measuring pauses
 - **units where pauses seldom occur** are seen as very cohesive: Possible enjambment contexts
- Breaking up such units can result in a noticeable effect

Quilis' enjambment typology

- **Tmesis:**
Breaks up a word
(1) The four eng-
ineers
Wore orange
brassieres
- **Phrase-bounded:**¹ A phrase gets split
(2) Listen! you hear the grating **roar**
Of pebbles with the waves draw back, and fling
- **Cross-clause:**² Between a relative pronoun
and its antecedent
(3) And time for all the works and days of **hands**
That lift and drop a question on your plate

¹encabalgamiento *sirremático* ²encabalgamiento *oracional* 17

Enjambment context examples

Some Phrase-bounded types

Noun or Adj Phrases

- determiner + noun
- noun or adj + prepositional complement
- adj + noun
- adj + adverb

Verb Phrases

- verb + adverb
- clitic + verb
- complex verb tenses
- verb + prepositional argument (Spanish *suplemento*)

COMPLETE TYPOLOGY

- <https://sites.google.com/site/spanishenjambment>

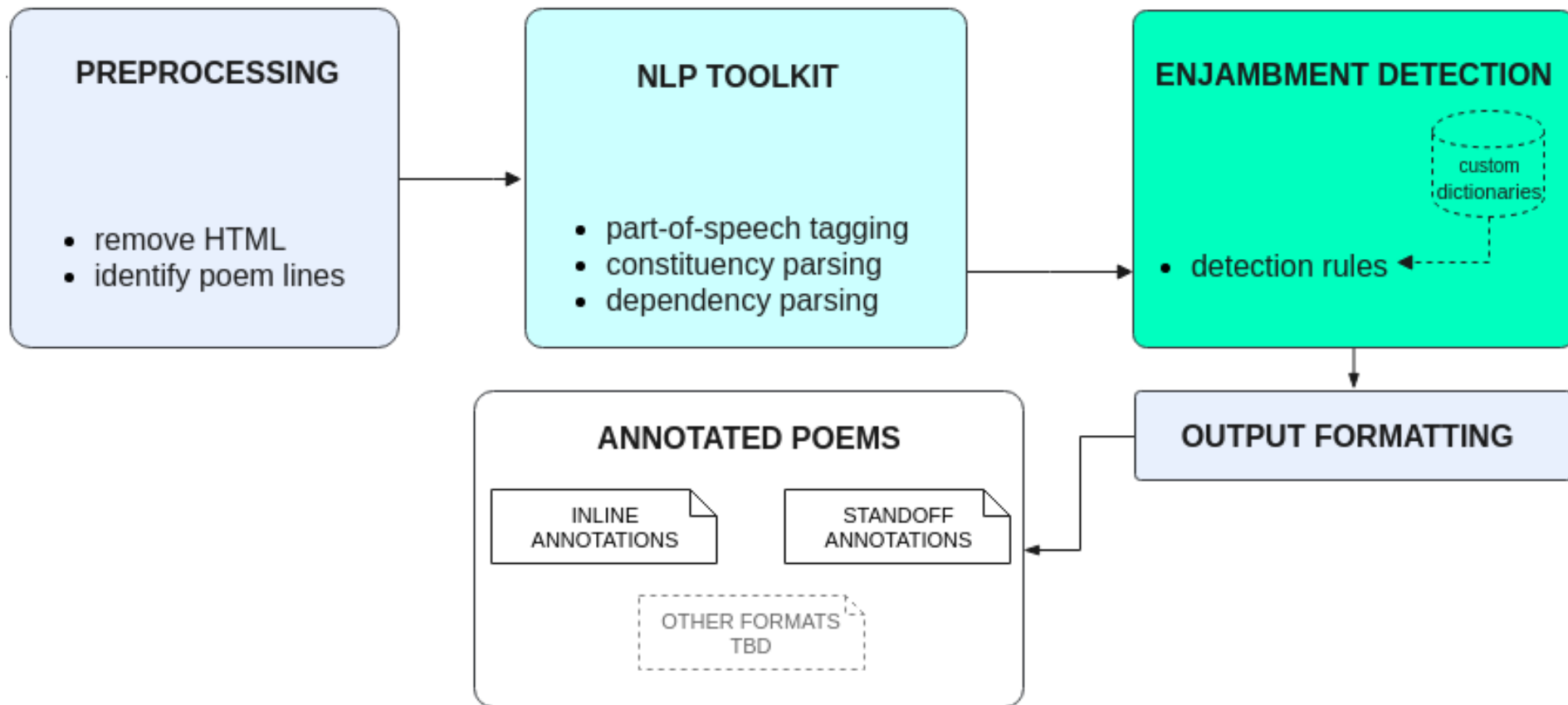
Automatic enjambment detection

- **Quilis**' definition, formulated in terms of
 - Part-of-speech sequences
 - Syntactic constituents
- **Spang** (1983) noted that breaking up a subject or object from its verb-head can also feel unnatural (*enlace* or *expansion*)
 - Syntactic dependencies
- NLP pipeline provides basis for detecting enjambment defined with such criteria

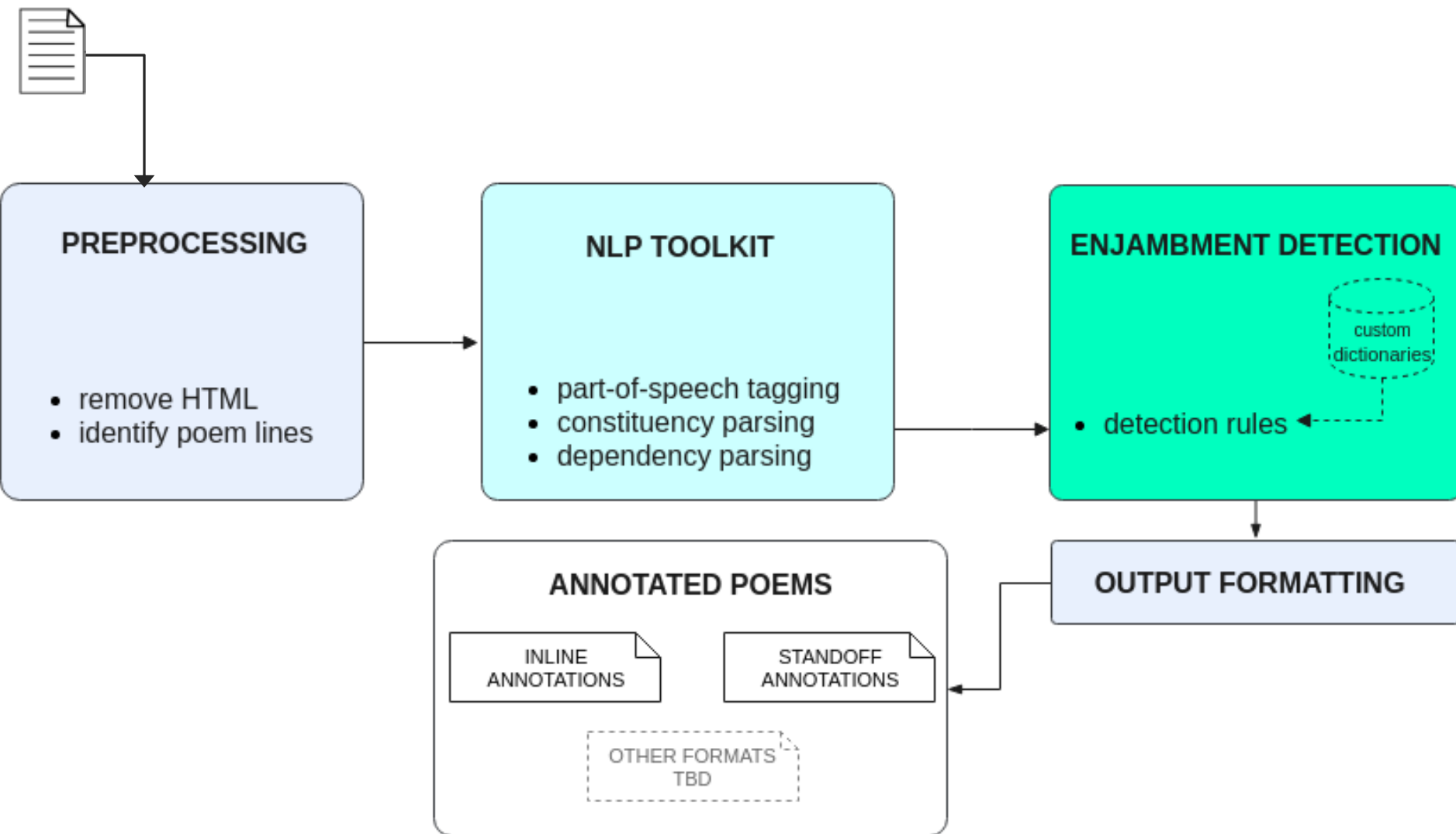
Talk Outline

1. Problem: Defining enjambment
2. Enjambment in Spanish
- 3. Enjambment detection system**
 - a. System description**
 - b. Evaluation: Corpus and results
4. Relevance for literary scholarship
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
5. Conclusion and outlook

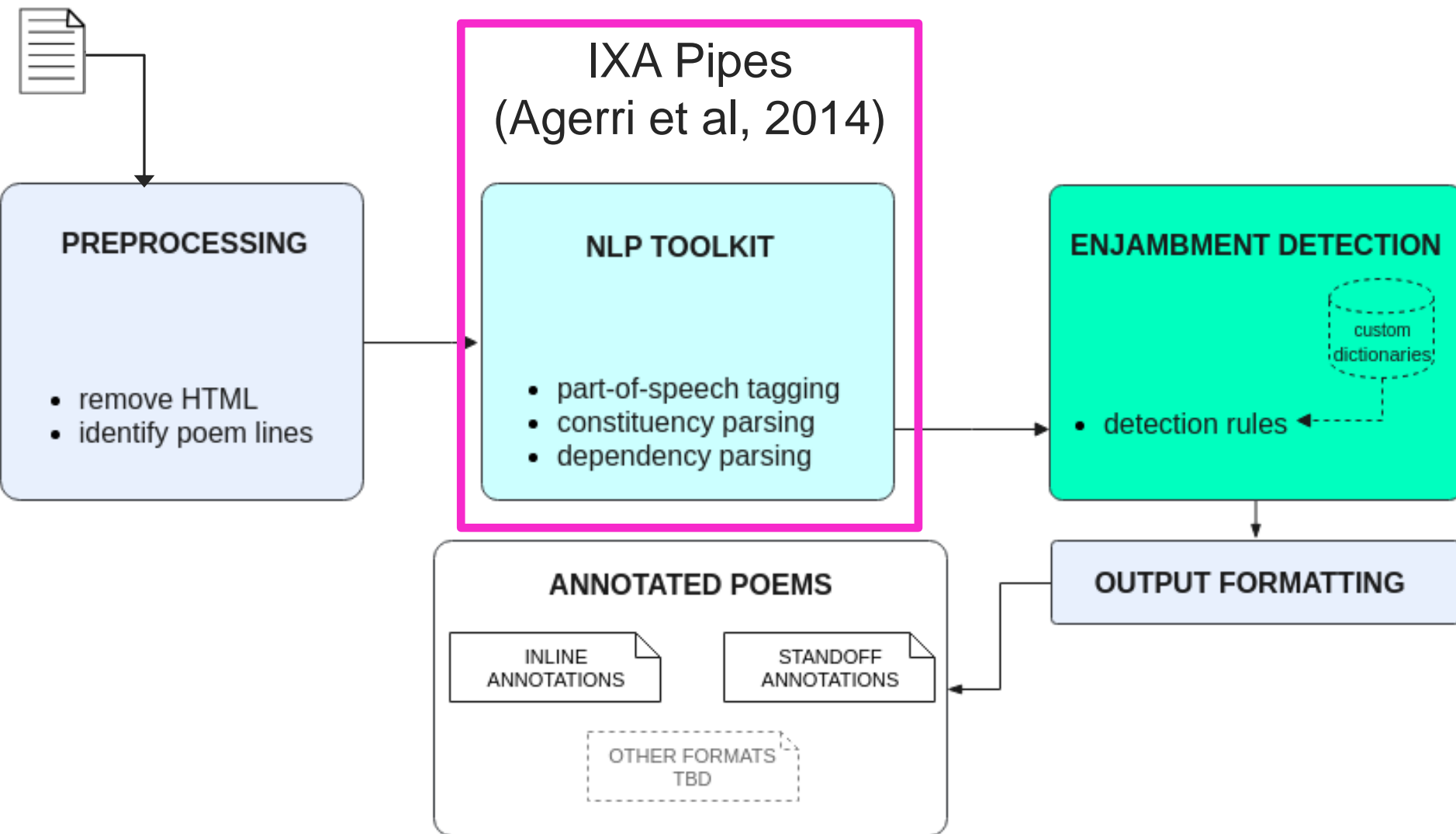
Enjambement detection system



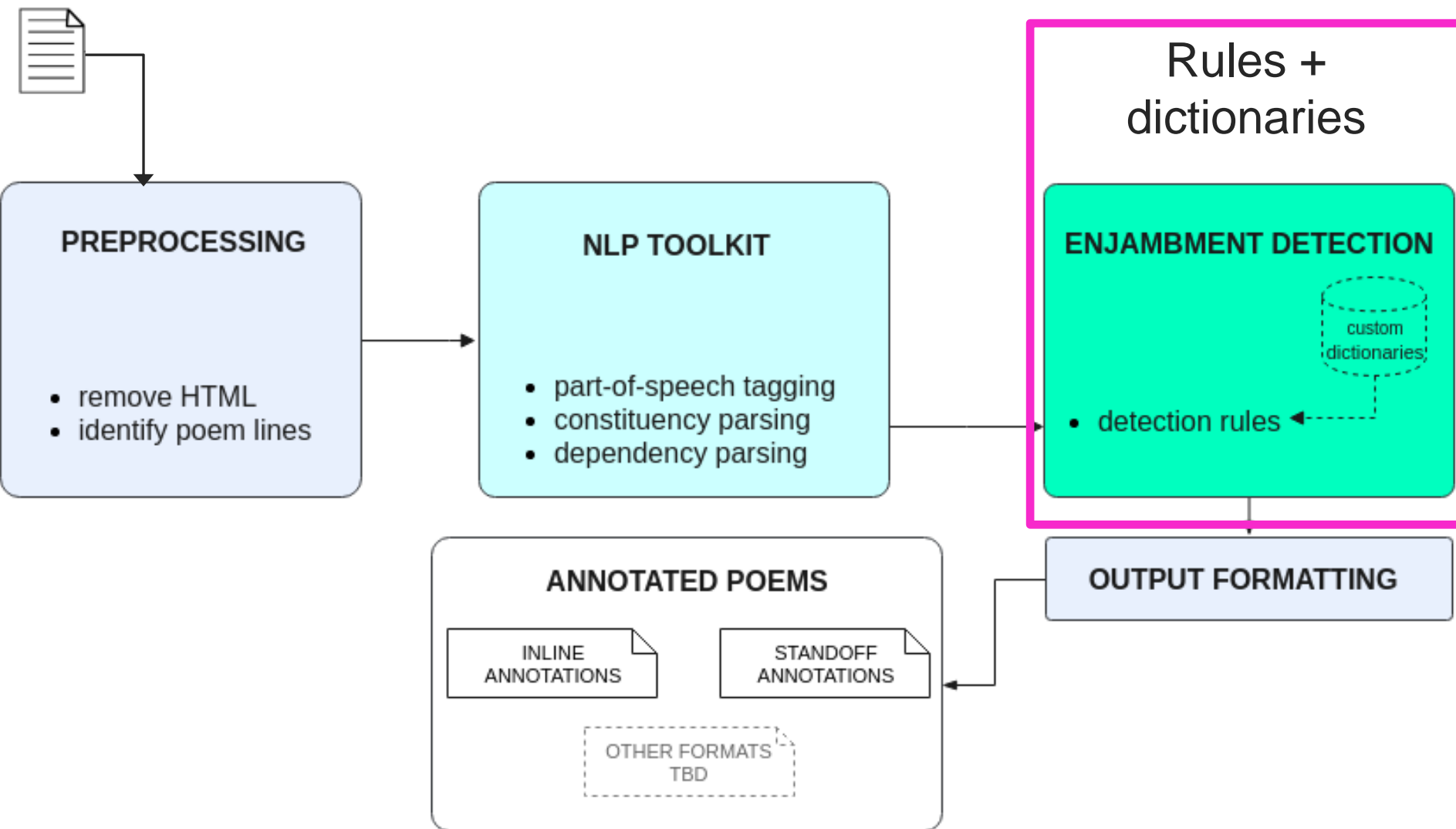
Enjambement detection system



Enjambement detection system



Enjambement detection system



Rule characteristics

- Approx. 30 rules
- Shallow rules based on POS sequences
- Such rules can also be lexically restricted to specific lemmas or word-forms
- Dependency-based
- Constituent-based

Rule characteristics

- Shallow rules based on POS sequences
 - 1 if, across a line-boundary, find *adj, noun, prep_de* (in any order)
 - 2 add type *pb_noun_prep* to the line-pair

Rule characteristics

- POS-based rules, but lexically restricted to specific lemmas or word-forms
- 1 if, immediately across a line-boundary, can match
 - **verb-lemma** from custom dictionary AND
 - the **preposition** whereby the matched verb can governs prepositional complements
 - 2 add **pb_verb_cprep** to the line-pair

Rule characteristics

- Dependency-based rules
 - 1 If, on adjacent lines, find a **verb** and its **subject**
 - 2 add **ex_subj_verb** to the line-pair
 - 1 If, on adjacent lines, find a verb and its object
 - 2 add **ex_dobj_verb** to the line-pair

Rule application example

- (4) 12 Si tú, Señor, le compadeces, cuida
13 de hacerle amar tu hora, la de verte,
14 si esperada quizás, siempre temida.

12 If you, Lord, take pity on him [my spirit], beware
13 of making him love your hour, that of seeing you,
14 which, while expected, is always feared

Amos de Escalante (1831-1901), “Al Cristo de mi cabecera”

Rule application example

- 12 Si tú, Señor, le compadeces, **cuida**
if you, Lord, him pity, beware
- 13 **de** hacerle amar tu hora, la de verte,
of make-him love your hour, that of seeing-you
- 14 si esperada quizás, siempre temida.
while expected perhaps, always feared

Rule application example

POS-TAGGING

12 {Si C} {tú Q} {, O} {Señor O} {, O} {le Q} {compadeces N} {, O} **{cuida V}**

13 **{de P}** {hacerle V} {amar V} {tu D} {hora N} {, O} {la D} {de P} {verte V} {, O}

14 {si C} {esperada G} {quizás A} {, O} {siempre A} {temida G} {. O}

Rule application example

POS-TAGGING

12 {Si C} {tú Q} {, O} {Señor O} {, O} {le Q} {compadeces N} {, O} **{cuida V}**
13 **{de P}** {hacerle V} {amar V} {tu D} {hora N} {, O} {la D} {de P} {verte V} {, O}
14 {si C} {esperada G} {quizás A} {, O} {siempre A} {temida G} {. O}

DEPENDENCIES

<deps>

[...]

<!--creg(cuida, de)-->

<dep from="t92" to="t93" rfunc="creg" />

[...]

</deps>

Rule application example

POS-TAGGING

12 {Si C} {tú Q} {, O} {Señor O} {, O} {le Q} {compadeces N} {, O} {**cuida V**}
13 {**de P**} {hacerle V} {amar V} {tu D} {hora N} {, O} {la D} {de P} {verte V} {, O}
14 {si C} {esperada G} {quizás A} {, O} {siempre A} {temida G} {, O}

DEPENDENCIES

<deps>

[...]

<!--creg(cuida, de)-->

<dep from="t92" to="t93" rfunc="creg" />

[...]

</deps>

RULE

- 1 if line n contains verb **v**
- 2 if line $n+1$ contains argument **arg** governed by v via a **creg** dependency
- 3 assign **pb_verb_cprep** to lines $\langle n, n+1 \rangle$

Rule application example

POS-TAGGING

12 {Si C} {tú Q} {, O} {Señor O} {, O} {le Q} {compadeces N} {, O} {**cuida V**}
13 {**de P**} {hacerle V} {amar V} {tu D} {hora N} {, O} {la D} {de P} {verte V} {, O}
14 {si C} {esperada G} {quizás A} {, O} {siempre A} {temida G} {, O}

DEPENDENCIES

<deps>

[...]

<!--creg(cuida, de)-->

<dep from="t92" to="t93" rfunc="creg" />

[...]

</deps>

RULE

- 1 if line n contains verb **v**
- 2 if line $n+1$ contains argument **arg** governed by v via a **creg** dependency
- 3 assign **pb_verb_cprep** to lines $\langle n, n+1 \rangle$

Output formats

- Standoff format for automatic evaluation

AuthorID	PoemID	Start	End	Type
269	0872	12	13	pb_verb_cprep

- Inline format with POS tags and term-IDs for manual inspection

Talk Outline

1. Problem: Defining enjambment
2. Enjambment in Spanish
- 3. Enjambment detection system**
 - a. System description
 - b. Evaluation: Corpus and results**
4. Relevance for literary scholarship
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
5. Conclusion and outlook

System Evaluation: Corpus

- Reference **corpus**
 - 1,000 lines of 20th century poems
 - 100 sonnets (1,400 lines) from the 15th to 19th centuries (30% from 19th)
 - Covers a variety of enjambment types
- Interannotator **agreement?**
 - 50% of test-items annotated by two people
 - Overlap for line-pairs with enjambment: 75.4%
 - Overlap for line-pairs without: 95.7%

System Evaluation: Tasks

- **Untyped** match (detection)
 - Line-pair positions proposed by system must match the reference
- **Typed** match (classification)
 - Line-pair positions and the type proposed by the system must match the reference

System Evaluation: Results

		N	P	R	F1
20 th	untyped	277	84.19	88.45	86.27
	typed	277	78.50	83.03	80.70

		N	P	R	F1
Dia-chronic	untyped	260	74.18	87.64	80.35
	typed	260	61.24	72.31	66.31

Number of line-pairs, Precision, Recall, F1

System Evaluation: Results

		N	P	R	F1
20 th	untyped	277	84.19	88.45	86.27
	typed	277	78.50	83.03	80.70

		N	P	R	F1
Dia-chronic	untyped	260	74.18	87.64	80.35
	typed	260	61.24	72.31	66.31

Number of line-pairs, Precision, Recall, F1

20th century: Typed match

Types	N	P	R	F1
<i>Phrase-bounded (all types)</i>	175	83.32	82.26	83.33
adj_adv	1	100	100	100
adj_noun	54	95.74	83.33	89.11
adj_prep	11	63.64	63.64	63.64
adv_prep	2	0	0	0
noun_prep	85	81.11	85.88	83.43
relword	2	100	100	100
verb_adv	7	77.78	100	87.50
verb_cprep	2	33.33	50.00	40.00
verb_chain	10	100	100	100
<i>Cross-clause</i>	31	100	93.55	96.67
<i>Expansions (by Spang)</i>	71	61.54	78.87	69.14
dobj_verb	39	66.67	87.18	75.56
subj_verb	32	55.00	68.75	61.11

20th century: Typed match

Types	N	P	R	F1
<i>Phrase-bounded (all types)</i>	175	83.32	82.26	83.33
adj_adv	1	100	100	100
adj_noun	54	95.74	83.33	89.11
adj_prep	11	63.64	63.64	63.64
adv_prep	2	0	0	0
noun_prep	85	81.11	85.88	83.43
relword	2	100	100	100
verb_adv	7	77.78	100	87.50
verb_cprep	2	33.33	50.00	40.00
verb_chain	10	100	100	100
<i>Cross-clause</i>	31	100	93.55	96.67
<i>Expansions (by Spang)</i>	71	61.54	78.87	69.14
dobj_verb	39	66.67	87.18	75.56
subj_verb	32	55.00	68.75	61.11

Diachronic corpus: Typed match

Type	N	P	R	F1
<i>Phrase-bounded (all types)</i>	104	66.19	88.46	75.72
adj_adv	2	100	50.00	66.67
adj_noun	29	54.55	82.76	65.75
adj_prep	14	58.82	71.43	64.52
noun_prep	39	55.36	79.49	65.26
relword	1	100	100	100
verb_adv	5	50.00	100	66.67
verb_cprep	9	83.33	55.56	66.67
verb_chain	5	100	80.00	88.89
<i>Cross-clause</i>	23	76.00	82.61	79.17
<i>Expansion (by Spang)</i>	133	61.54	66.17	63.77
dobj_verb	65	60.00	69.23	64.29
subj_verb	68	63.24	63.24	63.24

Diachronic corpus: Typed match

Type	N	P	R	F1
<i>Phrase-bounded (all types)</i>	104	66.19	88.46	75.72
adj_adv	2	100	50.00	66.67
adj_noun	29	54.55	82.76	65.75
adj_prep	14	58.82	71.43	64.52
noun_prep	39	55.36	79.49	65.26
relword	1	100	100	100
verb_adv	5	50.00	100	66.67
verb_cprep	9	83.33	55.56	66.67
verb_chain	5	100	80.00	88.89
<i>Cross-clause</i>	23	76.00	82.61	79.17
<i>Expansion (by Spang)</i>	133	61.54	66.17	63.77
dobj_verb	65	60.00	69.23	64.29
subj_verb	68	63.24	63.24	63.24

Error analysis

ko

ok

rule: pp05



20th century

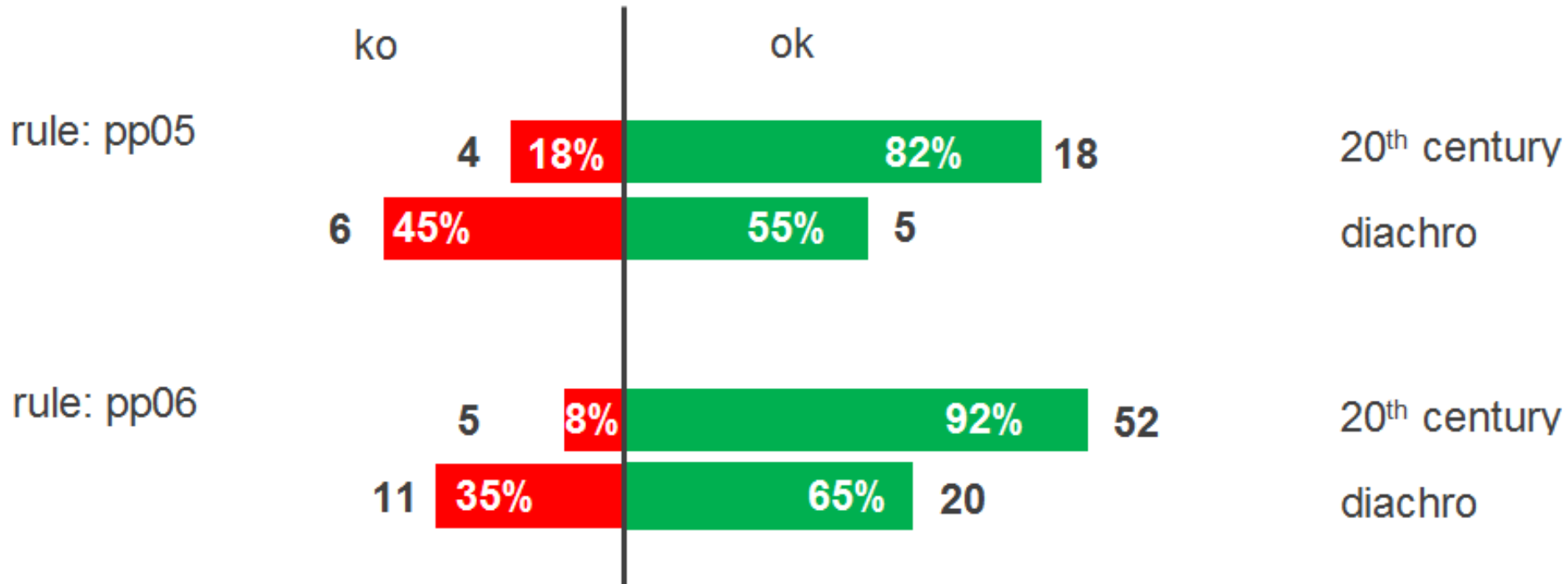
rule: pp06



20th century

rule_id	pos_sequence	position	enjambment_type
pp05	adj, noun, prep_de (in any order)	immediately across line	pb_noun_prep
pp06	noun, prep_de	boundary	

Error analysis



rule_id	pos_sequence	position	enjambment_type
pp05	adj, noun, prep_de (in any order)	immediately across line	pb_noun_prep
pp06	noun, prep_de	boundary	

Error sources

- Different error patterns in 20th century vs. diachronic corpus
- **20th century**: Besides rule errors, common NLP errors were:
 - Prepositional Phrase attachment
 - Confusion between subject and direct object
- **Diachronic**: Larger variety of error types, e.g.
 - POS-tagging errors
 - Rules overapply given hyperbaton

Error analysis: Conclusions

- Currently our rules exploit one source of information each:
 - POS sequence (plus lexical restrictions or not)
 - Dependency
 - Constituents
- Binary evaluation: Rule conditions met or not
- Manual error analysis showed that those information sources complement each other
- Combining them via machine learning may be better option (manual annotation ongoing)

Talk Outline

1. Problem: Defining enjambment
2. Enjambment in Spanish
3. Enjambment detection system
 - a. System description
 - b. Evaluation: Corpus and results
- 4. Relevance for literary scholarship**
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
5. Conclusion and outlook

Scholarly results

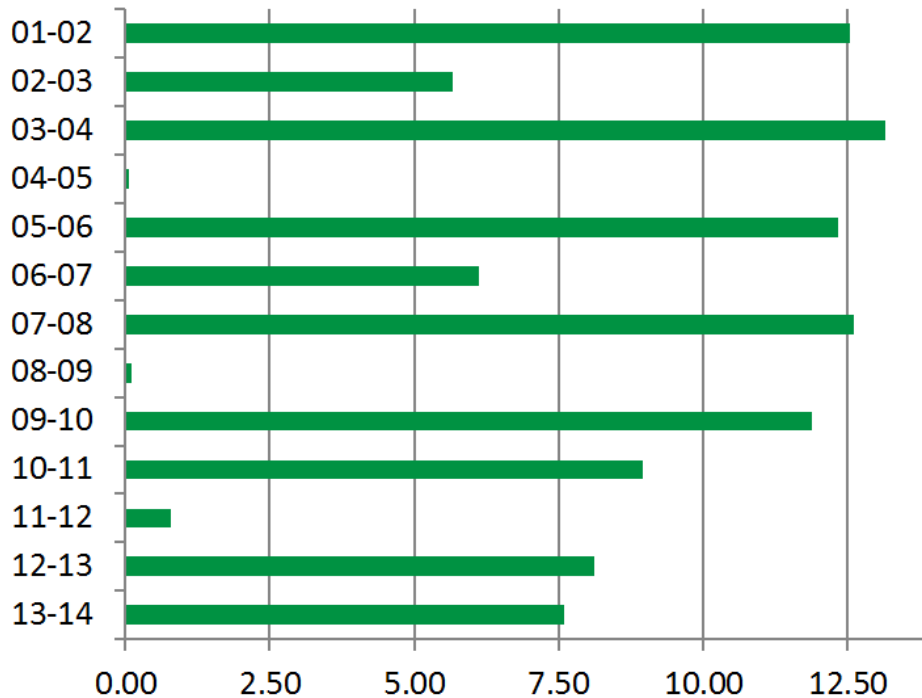
- Two goals
 - Quantitative results on the use of enjambment across genres or periods
 - Providing evidence to assess current characterizations of enjambment

Large diachronic corpus

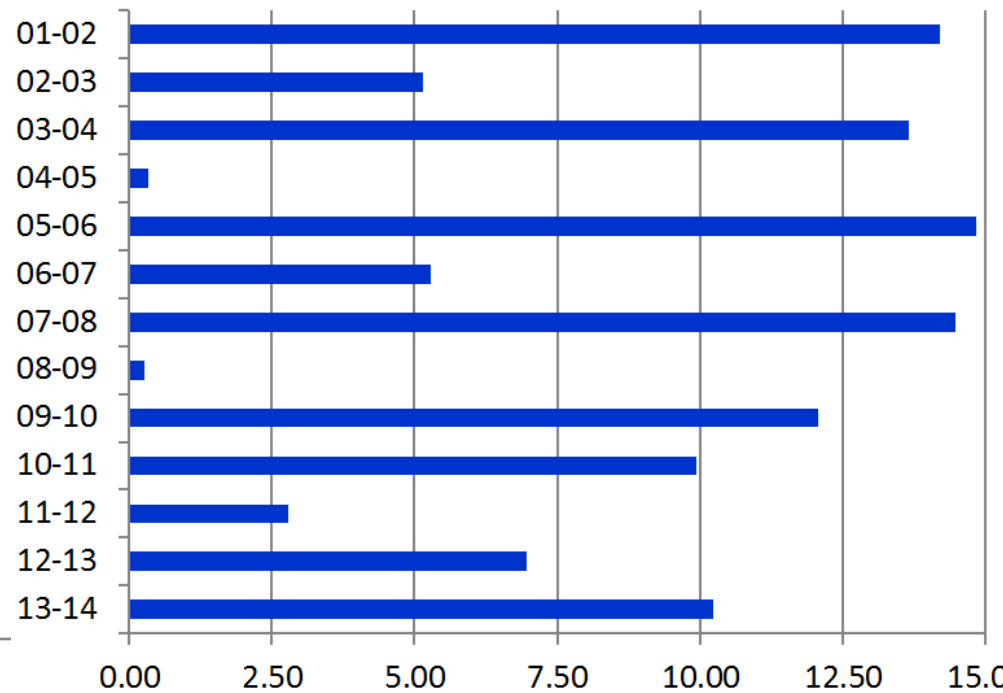
- Approx. 3750 sonnets
- Two periods
 - 15th to 17th centuries (ca. 1,000 sonnets)
 - 19th century (ca. 1,750 sonnets)
- Covers well-known and minor authors
- Created from online HTML sources
(anthologies by García González, 2006 at Biblioteca Virtual Cervantes)

Quantitative analysis per period: Distribution of enjambment per line-position

15th to 17th centuries



19th century



Percentage of enjambments per line-pair in sonnets

ciencias de las Juperiores. no dexare de traer esta aduersion, pues se ofrece lugar para ello; que cortar el verso en el Soneto, como,

**Quien me dixera, cuando en las passadas
oras —**

no es vicio si no virtud, i uno de los caminos principales para alcanzar l' alteza

alteza i hermosura del estilo; como en el Eroto latino, que romper el verso es grandeza del modo de dezir. refiero esto porque se persuaden algunos, que nunca dizen mejor, que cuando siempre acaban la sentencia con la rima. i óso afirmar, que ninguna mayor falta se puede casi hallar en el soneto; que terminar los versos deste modo. porque aunque sean compuestos de letras sonantes, i de silabas llenas casi todas., parecen de muy umilde estilo i simplicidad, no por flaqueza i desmayo de letras, si no por sola esta igual manera de passo, no apartando algun verso; que siendo todo entero a acabarse en su fin, no puede tener alguna cumplida gravedad, ni alteza, ni hermosura de estilo; si bien concurriessen todas las otras partes. Pero cuando quiere alguno acompañar el estilo confor=

Redefinition: Enjambment and hyperbaton

(5) Jauregui, _Juan_de ~~ Condena_al_fabricar_soberbios_palacios

01 ¡Ay, de cuán poco sirve al arrogante

02 el edificio que, soberbio empina

03 sobre pilastras de Tenaro, y **fina**

fine

04 **de mármol piedra**, y color cambiante!

of marble stone

- **noun** + **adjective**
- Without hyperbaton: **fina piedra de mármol**
- A pause between **fina** and **de mármol piedra** would be very unnatural

Redefinition: Enjambment and hyperbaton

(6) Acosta, _Ignacio_María ~~ Al_plan_de_Matanzas

12 ilumina mi espíritu; y lloroso,

13 que eres **la tumba perdurable** pienso

the tomb enduring think-I

14 del pueblo antiguo que habitó en Yucayo.

of-the people ancient

- Enjambment **noun** + **prepositional complement**
- Normal word order: **la tumba perdurable del pueblo antiguo**
- Pause after the **verb-form pienso** would be natural.
Can we speak of enjambment?

Demo

<http://evi.linhd.uned.es/anja/wspindex/>

Demo

ANJA

**Automatic enNJambment
Analysis**

<http://evi.lindh.uned.es/anja/wspindex/>

Standoff

Inline

PosTags

FullNLP

Submit

Oscuro oboe de bruma, cómo sepulta el mar
tu solemne sonido que despierta a los muertos.
Aquí, en esta ladera que cubre el olivar,
sangre y labio repiten musicales conciertos.
La ladera y sus soles que maduran los vinos,
la tensión del azul volcado de los cielos,
armonía y vacío en espacios divinos,
horror al más allá tras las costas de Delos.
Perdición, extravío de las horas oscuras
del corazón que sueña, en la luz, otros mundos,
mientras la boca va desgranando las puras
notas de arte mortal en los hondos, inmundos
socavones del mal. Pues ¿por qué se da guerra
junto al amor y por qué la voz de la ebriedad
y el dolor infinito van girando en la tierra?
Un año más se funden misterio y soledad.
Y, sin embargo, tú, estrella de la tarde,
ves llegar el final, los bordes del ocaso.
¿Qué secretos oculta este cosmos que arde
en la muerte y qué nos reserva el ocaso?
Mas, en el hondo instante, la música revela
la inmensidad del orbe, la dimensión del ser.
Un aroma de azahar la angustia nos desvela.
Horror y sed de dioses tras otro atardecer.
Oscuro oboe de bruma, entreabre las venas
del mundo en esta paz y arrasa la Historia.
Vida y muerte se acercan como olas serenas
al corazón que ahuyenta, soñando, la memoria.

Submit

Standoff

Inline

PosTags

FullNLP

Start	End	Enjambment Type
1	2	enlace_od_verb
9	10	sirrem_noun_prep
12	13	sirrem_adj_noun
13	14	sirrem_adj_noun
14	15	enlace_subj_verb
21	22	enlace_od_verb
25	26	sirrem_noun_prep

Standoff

Inline

PosTags

FullNLP

#	Text	Position	Enjambment Type
1	{Oscuro D} {oboe N} {de P} {bruma N} {, O} {cómo Q} {sepulta V} {el D} {mar N}	B	enlace_od_verb
2	{tu D} {solemne G} {sonido N} {que Q} {despierta V} {a P} {los D} {muertos N} {, O}	I	enlace_od_verb
3	{Aquí A} {, O} {en P} {esta D} {ladera N} {que Q} {cubre V} {el D} {olivar N} {, O}	O	
4	{sangre N} {y C} {labio N} {repiten V} {musicales G} {conciertos N} {, O}	O	
5	{La D} {ladera N} {y C} {sus D} {soles N} {que Q} {maduran V} {los D} {vinos N} {, O}	O	
6	{la D} {tensión N} {del P} {azul N} {volcado G} {de P} {los D} {cielos N} {, O}	O	
7	{armonía N} {y C} {vacío N} {en P} {espacios N} {divinos G} {, O}	O	
8	{horror N} {al P} {más A} {allá A} {tras P} {las D} {costas N} {de P} {Delos R} {, O}	O	
9	{Perdición N} {, O} {extravío N} {de P} {las D} {horas N} {oscuras G}	B	sirrem_noun_prep
10	{del P} {corazón N} {que Q} {sueña V} {, O} {en P} {la D} {luz N} {, O} {otros D} {mundos N} {, O}	I	sirrem_noun_prep
11	{mientras C} {la D} {boca N} {va V} {desgranando V} {las D} {puras N}	O	
12	{notas N} {de P} {arte N} {mortal G} {en P} {los D} {hondos N} {, O} {inmundos G}	B	sirrem_adj_noun
13	{socavones N} {del P} {mal N} {, O} {Pues C} {¿ O} {por P} {qué Q} {se Q} {da V} {guerra N}	IB	sirrem_adj_noun sirrem_adj_noun
14	{junto G} {al P} {amor N} {y C} {por P} {qué Q} {la D} {voz N} {de P} {la D} {ebriedad N}	IB	sirrem_adj_noun enlace_subj_verb
15	{y C} {el D} {dolor N} {infinito G} {van V} {sairando V} {en P} {la D} {tierra N} {¿ O}	I	enlace_subj_verb

#	Text
1	{Oscuro D t1} {oboe N t2} {de P t3} {bruma N t4} {, O t5} {cómo Q t6} {sepulta V t7} {el D t8} {mar N t9}
2	{tu D t10} {solemne G t11} {sonido N t12} {que Q t13} {despierta V t14} {a P t15} {los D t16} {muertos N t17} {, O t18}
3	{Aquí A t19} {, O t20} {en P t21} {esta D t22} {ladera N t23} {que Q t24} {cubre V t25} {el D t26} {olivar N t27} {, O t28}
4	{sangre N t29} {y C t30} {labio N t31} {repiten V t32} {musicales G t33} {conciertos N t34} {, O t35}
5	{La D t36} {ladera N t37} {y C t38} {sus D t39} {soles N t40} {que Q t41} {maduran V t42} {los D t43} {vinos N t44} {, O t45}
6	{la D t46} {tensión N t47} {del P t48} {azul N t49} {volcado G t50} {de P t51} {los D t52} {cielos N t53} {, O t54}
7	{armonía N t55} {y C t56} {vacío N t57} {en P t58} {espacios N t59} {divinos G t60} {, O t61}
8	{horror N t62} {al P t63} {más A t64} {allá A t65} {tras P t66} {las D t67} {costas N t68} {de P t69} {Delos R t70} {, O t71}
9	{Perdición N t72} {, O t73} {extravío N t74} {de P t75} {las D t76} {horas N t77} {oscuras G t78}
10	{del P t79} {corazón N t80} {que Q t81} {sueña V t82} {, O t83} {en P t84} {la D t85} {luz N t86} {, O t87} {otros D t88} {mundos N t89} {, O t90}
11	{mientras C t91} {la D t92} {boca N t93} {va V t94} {desgranando V t95} {las D t96} {puras N t97}
12	{notas N t98} {de P t99} {arte N t100} {mortal G t101} {en P t102} {los D t103} {hondos N t104} {, O t105} {inmundos G t106}
13	{socavones N t107} {del P t108} {mal N t109} {, O t110} {Pues C t111} {¿ O t112} {por P t113} {qué Q t114} {se Q t115} {da V t116} {guerra N t117}
14	{junto G t118} {al P t119} {amor N t120} {y C t121} {por P t122} {qué Q t123} {la D t124} {voz N t125} {de P t126} {la D t127} {ebriedad N t128}
15	{y C t129} {el D t130} {dolor N t131} {infinito G t132} {van V t133} {girando V t134} {en P t135} {la D t136} {tierra N t137} {? O t138}
16	{Un D t139} {año N t140} {más A t141} {se O t142} {funden V t143} {misterio N t144} {y C t145} {soledad N t146} {, O t147}



```

1. <?xml version="1.0" encoding="UTF-8"?>
2. <NAF xml:lang="es" version="v1.naf">
3.   <nafHeader>
4.     <linguisticProcessors layer="text">
5.       <lp name="ixa-pipe-tok-es" beginTimestamp="2017-10-02T10:29:52+0200" endTimestamp="2017-10-02T10:29:52+0200" v
6.     </linguisticProcessors>
7.     <linguisticProcessors layer="terms">
8.       <lp name="ixa-pipe-pos-es-pos-perceptron-autodict01-ancora-2.0" beginTimestamp="2017-10-02T10:29:52+0200" endT
9.     </linguisticProcessors>
10.    <linguisticProcessors layer="deps">
11.      <lp name="ixa-pipe-srl-es" beginTimestamp="2017-10-02T10:29:52+0200" endTimestamp="2017-10-02T10:29:53+0200" v
12.    </linguisticProcessors>
13.  </nafHeader>
14.  <text>
15.    <wf id="w1" offset="0" length="6" sent="1" para="1">Oscuro</wf>
16.    <wf id="w2" offset="7" length="4" sent="1" para="1">oboe</wf>
17.    <wf id="w3" offset="12" length="2" sent="1" para="1">de</wf>
18.    <wf id="w4" offset="15" length="5" sent="1" para="1">bruma</wf>
19.    <wf id="w5" offset="20" length="1" sent="1" para="1">,</wf>
20.    <wf id="w6" offset="22" length="4" sent="1" para="1">cómo</wf>
21.    <wf id="w7" offset="27" length="7" sent="1" para="1">sepulta</wf>
22.    <wf id="w8" offset="35" length="2" sent="1" para="1">el</wf>
23.    <wf id="w9" offset="38" length="3" sent="1" para="1">mar</wf>
24.    <wf id="w10" offset="42" length="2" sent="1" para="1">tu</wf>
25.    <wf id="w11" offset="45" length="7" sent="1" para="1">solemne</wf>
26.    <wf id="w12" offset="53" length="6" sent="1" para="1">sonido</wf>
27.    <wf id="w13" offset="60" length="3" sent="1" para="1">que</wf>
28.    <wf id="w14" offset="64" length="9" sent="1" para="1">despierta</wf>
29.    <wf id="w15" offset="74" length="1" sent="1" para="1">a</wf>
30.    <wf id="w16" offset="76" length="3" sent="1" para="1">los</wf>

```

Talk Outline

1. Problem: Defining enjambment
2. Enjambment in Spanish
3. Enjambment detection system
 - a. System description
 - b. Evaluation: Corpus and results
- 4. Relevance for literary scholarship**
 - a. Quantitative studies on a large corpus
 - b. Refining definitions
- 5. Conclusion and outlook**

Conclusion and Outlook

- Rule-based system to detect and classify enjambment (and “expansion”) in Spanish based on POS or dependencies or constituents
- Error analysis suggests combining those sources to improve results
 - Manual annotation ongoing: supervised learning?
- Adaptation to older language varieties and poetic diction would be advisable
- Literary results on a sonnet corpus covering four centuries suggest system’s usefulness

Sources for the examples cited

- (1) Tmesis example: Attributed to Willard Espy (20th century)
- (2) “Primero de un año” (“The first day of some year”). Gloria Fuertes (1980), cited in García-Page (1991)
- (3) “Dover Beach”, Matthew Arnold (1867)
- (4) “The Love Song of J. Alfred Prufrock”, T. S. Eliot (1915)
- (5) “Al Cristo de mi cabecera” (“To the crucifix in my room”). Amos Escalante (19th century)
- (6) “Condena al fabricar soberbios palacios”. (“Condemnation upon building lavish palaces”). Juan de Jáuregui (16th-17th centuries)

Thanks!

PROJECT SITE

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

cimartinez@flog.uned.es

pablo.ruiz@linhd.uned.es

