

# Roman de Flamenca: Annotation Description

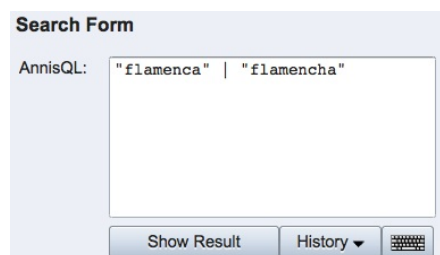
## 1. Introduction

1. Web database [http://nlp.indiana.edu/~obscrivn/F1\\_495.html](http://nlp.indiana.edu/~obscrivn/F1_495.html)
2. Corpus visual search <http://nlp.indiana.edu:8080/annis-gui-3.0.0/>
3. Select subcorpus (lines) from Corpus List: ex. 1\_992 or Align1\_922 (for parallel text)

## 2. Token Annotation - Word Search

- a.) Double quotation: “flamenca” (no capitalized words)  
Pipe sign (|) - one or another: “flamenca”|“flamencha”
- b.) Regular expression - spelling variation: /flamen.+/ (“.” equal any character and “+” equal one or more times). For more information on search with regular expressions (.\*+?), see Chapter 11 from *Corpus Linguistics and Linguistically Annotated Corpora* (Sandra Kübler and Heike Zinsmeister, 2014)

Search example in Text Query Window:



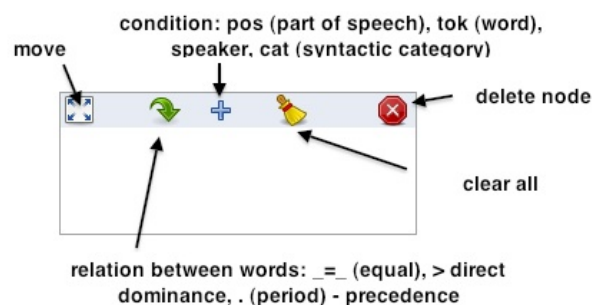
Search Form

AnnisQL: "flamenca" | "flamencha"

Show Result History

## 3. Morphological Annotation - Part of Speech Search

- a.) Double quotation: pos="NPRS" (searching for a proper singular noun - see Table 1)  
**pos** - Occitan, **epos** - English
- b.) Combination of two or more words (tags) - use Graphical Query Builder Panel:



Search example: verb pos="VJ" precedes pronoun pos="PRO"

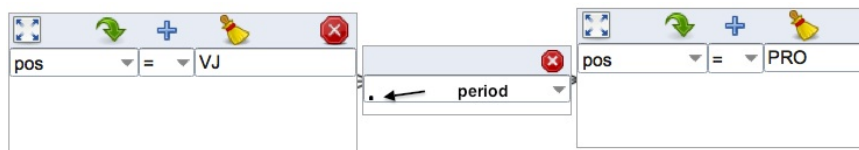


Table 1: Occitan Part-of-Speech tagset (adapted from MCVF corpus description, Martineau et al. 2007)

Tag	Definition	Tag	Definition
ADJ	adjective	MDPP	past participle of modal verb
ADJR	comparative form of adjective	MDX	infinitive of modal verb
ADV	adverb	NCPL	noun common plural
ADVR	comparative form of adverb	NCS	noun common singular
AG	gerundive of auxiliary 'to have'	NEG	negation
AJ	present of auxiliary 'to have'	NPRPL	noun proper plural
APP	past participle of auxiliary 'to have'	NPRS	noun proper singular
AX	infinitive of auxiliary 'to have'	NUM	numeral
CONJO	coordinative conjunction	P	preposition
CONJS	subordinate conjunction	PON	punctuation inside the clause
COMP	comparative adverb	PONFP	the end of the sentence
D	determiner (indefinite, definite, demonstrative)	PRO	pronoun
DAT	dative	Q	quantifier
DZ	possessive determiner	VG	gerundive of the main verb
EG	gerundive of auxiliary 'to be'	VJ	present of the main verb
EJ	present of auxiliary 'to be'	VPP	past participle of the main verb
EPP	past participle of auxiliary 'to be'	VX	infinitive of the main verb
EX	infinitive of auxiliary 'to be'	WADV	interr., rel. or excl. adverb
ITJ	interjection	WD	interr., rel. or excl. determiner
MDG	gerundive of modal verb	WPRO	interr., rel. or excl. pronoun
MDJ	present of modal verb	QUO	quotation mark

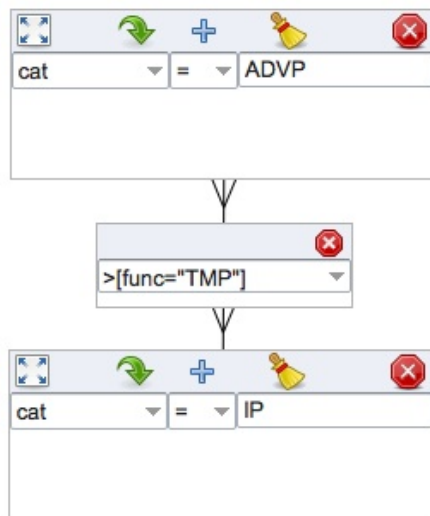
#### 4. Syntactic Annotation

Table 2: Occitan Syntactic Labels (adapted from MCVF corpus description, Martineau et al. 2007)

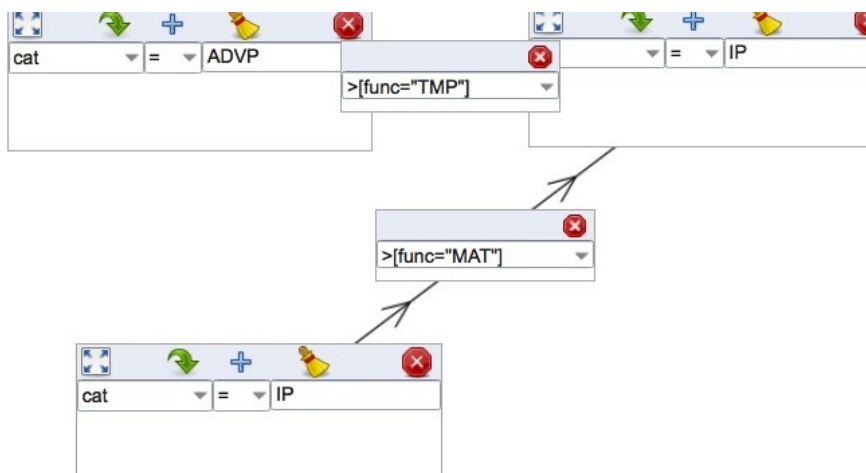
Labels	Definition	Label	Definition
ADJP	Adjectival Phrase	CP-THT	Complement Clause
ADVP	Adverbial Phrase	INTJ	Interjection
ADVP-LOC	Adverbial Locative Phrase	IP-IMP	Imperative Proposition
ADVP-TMP	Adverbial Temporal Phrase	IP-INF	Infinitival Proposition
CONJP	Conjunction	IP-MAT	Main Proposition
CP-ADV	Adverbial Clause	IP-PPL	Participial Proposition
CP-ADV-TMP	Temporal Clause	IP-SUB	Subordinate Proposition
CP-CAR	Prepositional Clause	NP-ACC	Direct Object
CP-CMP	Comparative Clause	NP-DTV	Indirect Object
CP-DEG	Degree Clause	NP-PRD	Predicative NP
CP-EXL	Exclamative Clause	NP-RFL	Reflexive NP
CP-FRL	Small Clause	NP-SBJ	Subject NP
CP-OPT	Optative Clause	NP-TMP	Temporal NP
CP-QUE	Interrogative Clause	PP	Prepositional Phrase
CP-REL	Relative Clause	PP-DIR	Prepositional Directional Phrase
PP-LOC	Prepositional Locative Phrase	QR	Quantifier Phrase
V	Verb		
-LFD	Left Dislocated Phrase	-PRN	Adjunct
-SPE	Direct Speech		

Search examples in Graphical Query Window:

1. Temporal adverbial phrases:



2. Temporal adverbial phrases in main clauses:



## 5. Discourse Annotation

Speakers: Archambaut, Flamenca, Guillem, King, Queen, MaleSpeaker, FemaleSpeaker, Author, narration.

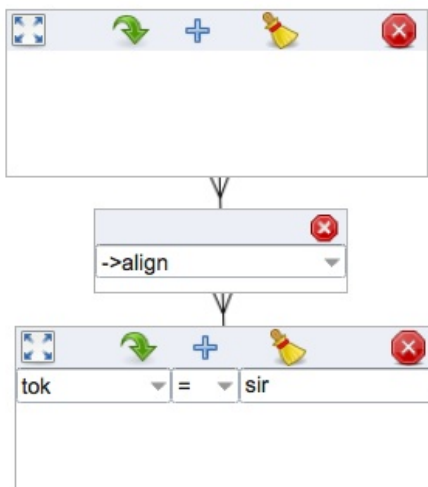
Text Window Query: **speaker**="Flamenca"

## 6. Parallel Alignment

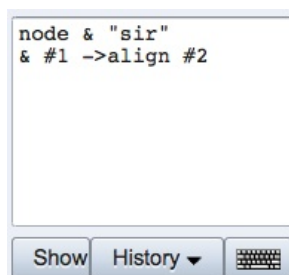
- Alignment is between nodes
- For English par of speech tag set, consult: <http://www.americannationalcorpus.org/OANC/penn.html>
- Example for Graphical Window - Search for English word "sir"  
Create a blank node for Occitan, create a node for English with token (tok) equal "sir" and connect

two nodes by ->align.

To connect two nodes, choose a green arrow on Occitan node and ”dock” it on English node:



- The same example in Text Window Search:



## 7. Lemma

Occitan sections provide a lemma search (dictionary form) for tokens. Ex. you can search for **lemma**="faire" and find all occurrences of this verb. Ex. *faria, fan, fai* etc.

## 8. Export Results

Go to **Export** window and select Exporter "TextExporter", click Perform Export and then Download. Left Context specifies how many tokens before and Right Context how many tokens after the search form.

## 9. Collaboration

This project is supported by passion and enthusiasm of the very small group. If you are interested in becoming part of this effort, or you have questions about search queries, or you have found errors in annotation, do not hesitate to contact us! (Olga Scrivner [obscrivn@indiana.edu](mailto:obscrivn@indiana.edu), Sandra Kübler [skuebler@indiana.edu](mailto:skuebler@indiana.edu), Barbara Vance [bvance@indiana.edu](mailto:bvance@indiana.edu))