

Análisis lingüístico automático

Escuela Nacional de Antropología e Historia (ENAH)
Agosto – diciembre de 2015

Tokenización (Tokenization)

- Separar un texto en palabras gráficas (tokens)
- Nombres propios
- Fechas
- Abreviaturas
- Términos multipalabra

Truncamiento (Stemming)

- Obtener la base de una palabra.
- Eliminar solo los afijos de flexión.

Lematización (Lemmatization)

- Obtener la forma canónica de un conjunto de variantes morfológicas.

Etiquetado de categorías gramaticales (Part-of-speech Tagging)

- Asignar clase y categoría gramatical a cada palabra del corpus.

Reconocimiento de entidades (Named Entity Recognition)

- Reconocer nombres de personas, organizaciones y lugares.

FreeLing Home Page

HAVE YOU GOT A FREELING ?

Menu

[Home](#)

[Features](#)

[Linguistic Data](#)

[Contributions](#)

[License](#)

[Requirements](#)

[Installing](#)

[Documentation](#)

[Contributing](#)

[Download](#)

[SVN](#)

[References](#)

[Links](#)

[On-line demo](#)

[Forum & FAQs](#)

FreeLing 3.1

AN OPEN SOURCE SUITE OF LANGUAGE ANALYZERS

Here you can find information about FreeLing, an open source language analysis tool suite, released under the [GNU GPL](#).

The FreeLing project was created and is currently leaded by [Lluís Padró](#) as a means to make available to the research community a tool suite for the [Natural Language Processing](#) group.

What's new in FreeLing 3.1:

- ▶ FreeLing 3.1 is thread-safe, allowing parallel processing (e.g. in webservice).
- ▶ Extended API to analysis modules.
- ▶ Improved API to manage and navigate resulting linguistic structures.
- ▶ New functionalities to deal with non-standard text: SED-based search of dictionary words similar to `g`.
- ▶ Wrapped regular expression module that enables to choose between `boost::regex` and `boost::xpres`.
- ▶ New languages: Up to PoS-tagging for French. Partial support for Czech and Slovenian.

FreeLing 3.1

AN OPEN-SOURCE SUITE OF LANGUAGE ANALYZERS

Write your sentences

```
Mi amigo Juan Mesa se mesa la
barba al lado de la mesa hoy
18 de septiembre.
```

Analysis options

- Multiword detection
- Number recognition
- Date/Time recognition
- Quantities, ratios, and percentages
- Named Entity detection
- Named Entity classification
- Phonetic encoding
- No sense annotation
- WN sense annotation: Frequency so
- WN sense annotation: PageRank so

Select language

Spanish ▾

Select output

Morphological Analysis ▾

Submit

Analysis Results

Sentence #1

Mi	amigo	Juan	Mesa	se	mesa	la
<i>mi</i> DP1CSS 0.999101	<i>amigo</i> NCMS000 0.893805	<i>juan</i> VMIP3P0 0.460049	<i>mesa</i> NCFS000 0.980583	<i>se</i> P00CN000 0.465639	<i>mesa</i> NCFS000 0.980583	<i>el</i> DA0FS0 0.972269
<i>mi</i> NCMS000 0.000899281	<i>amigo</i> AQ0MS0 0.0973451	<i>juan</i> VMII3P0 0.261756	<i>mesar</i> VMIP3S0 0.00970874	<i>se</i> P03CN000 0.465639	<i>mesar</i> VMIP3S0 0.00970874	<i>lo</i> PP3FSA00 0.0277039
	<i>amigar</i> VMIP1S0 0.00884956	<i>juan</i> RG 0.110705	<i>mesar</i> VMM02S0 0.00970874	<i>se</i> PP3CN000 0.0687218	<i>mesar</i> VMM02S0 0.00970874	<i>la</i> NCMS000 2.74025e-05

la	mesa	hoy	18	de	septiembre	.
<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>	<hr/>
<i>el</i>	<i>mesa</i>	<i>hoy</i>	<i>18</i>	<i>de</i>	<i>septiembre</i>	<i>.</i>
DA0FS0	NCFS000	RG	NCFS000	SPS00	NCMS000	Fp
0.972269	0.980583	0.883142	0.15945	0.999984	1	1
<hr/>	<hr/>	<hr/>	<hr/>	<hr/>		
<i>lo</i>	<i>mesar</i>	<i>hoy</i>	<i>18</i>	<i>de</i>		
PP3FSA00	VMIP3S0	NCMS000	NCMS000	NCFS000		
0.0277039	0.00970874	0.116858	0.15899	1.61912e-05		
<hr/>	<hr/>		<hr/>			
<i>la</i>	<i>mesar</i>		<i>18</i>			
NCMS000	VMM02S0		RG			
2.74025e-05	0.00970874		0.106203			
			<hr/>			

FreeLing 3.1

AN OPEN-SOURCE SUITE OF LANGUAGE ANALYZERS

Write your sentences

Mi amigo Juan Mesa se mesa la
barba al lado de la mesa hoy
18 de septiembre.

Analysis options

- Multiword detection
- Number recognition
- Date/Time recognition
- Quantities, ratios, and percentages
- Named Entity detection
- Named Entity classification
- Phonetic encoding
- No sense annotation
- WN sense annotation: Frequency sort
- WN sense annotation: PageRank sort

Select language

Spanish ▾

Select output

Morphological Analysis ▾

Submit

Analysis Results

Sentence #1

Mi	amigo	Juan_Mesa	se	mesa	la	barba
<i>mi</i> DP1CSS 0.999101	<i>amigo</i> NCMS000 0.893805	<i>juan_mesa</i> NP00000 1	<i>se</i> P00CN000 0.465639	<i>mesa</i> NCFS000 0.980583	<i>el</i> DA0FS0 0.972269	<i>barba</i> NCFS000 0.846947
<i>mi</i> NCMS000 0.000899281	<i>amigo</i> AQ0MS0 0.0973451		<i>se</i> P03CN000 0.465639	<i>mesar</i> VMIP3S0 0.00970874	<i>lo</i> PP3FSA00 0.0277039	<i>barbar</i> VMIP3S0 0.150743
	<i>amigar</i> VMIP1S0 0.00884956		<i>se</i> PP3CN000 0.0687218	<i>mesar</i> VMM02S0 0.00970874	<i>la</i> NCMS000 2.74025e-05	<i>barbar</i> VMM02S0 0.00231023

de	la	mesa	hoy	18_de_septiembre	.
<i>de</i> SPS00 0.999984	<i>el</i> DA0FS0 0.972269	<i>mesa</i> NCFS000 0.980583	<i>hoy</i> RG 0.883142	[?:18/9 /?:?.?:?] W 1	. Fp 1
<i>de</i> NCFS000 1.61912e-05	<i>lo</i> PP3FSA00 0.0277039	<i>mesar</i> VMIP3S0 0.00970874	<i>hoy</i> NCMS000 0.116858		
	<i>la</i> NCMS000 2.74025e-05	<i>mesar</i> VMM02S0 0.00970874			

Select language

Spanish

Select output

PoS Tagging

Submit

Analysis Results

Sentence #1

Mi	amigo	Juan_Mesa	se	mesa	la	barba	a
<i>mi</i>	<i>amigo</i>	<i>juan_mesa</i>	<i>se</i>	<i>mesar</i>	<i>el</i>	<i>barba</i>	<i>a</i>
DP1CSS	NCMS000	NP00000	P00CN000	VMIP3S0	DA0FS0	NCFS000	SPS00

el **lado** **de** **la** **mesa** **hoy** **18_de_septiembre** **.**
el *lado* *de* *el* *mesa* *hoy* *[?:18/9* *.*
DA0MS0 NCMS000 SPS00 DA0FS0 NCFS000 RG /?:?.?:?]? Fp
W

Análisis sintáctico superficial (Shallow parsing)

- Agrupar palabras en frases nominales, verbales y preposicionales .

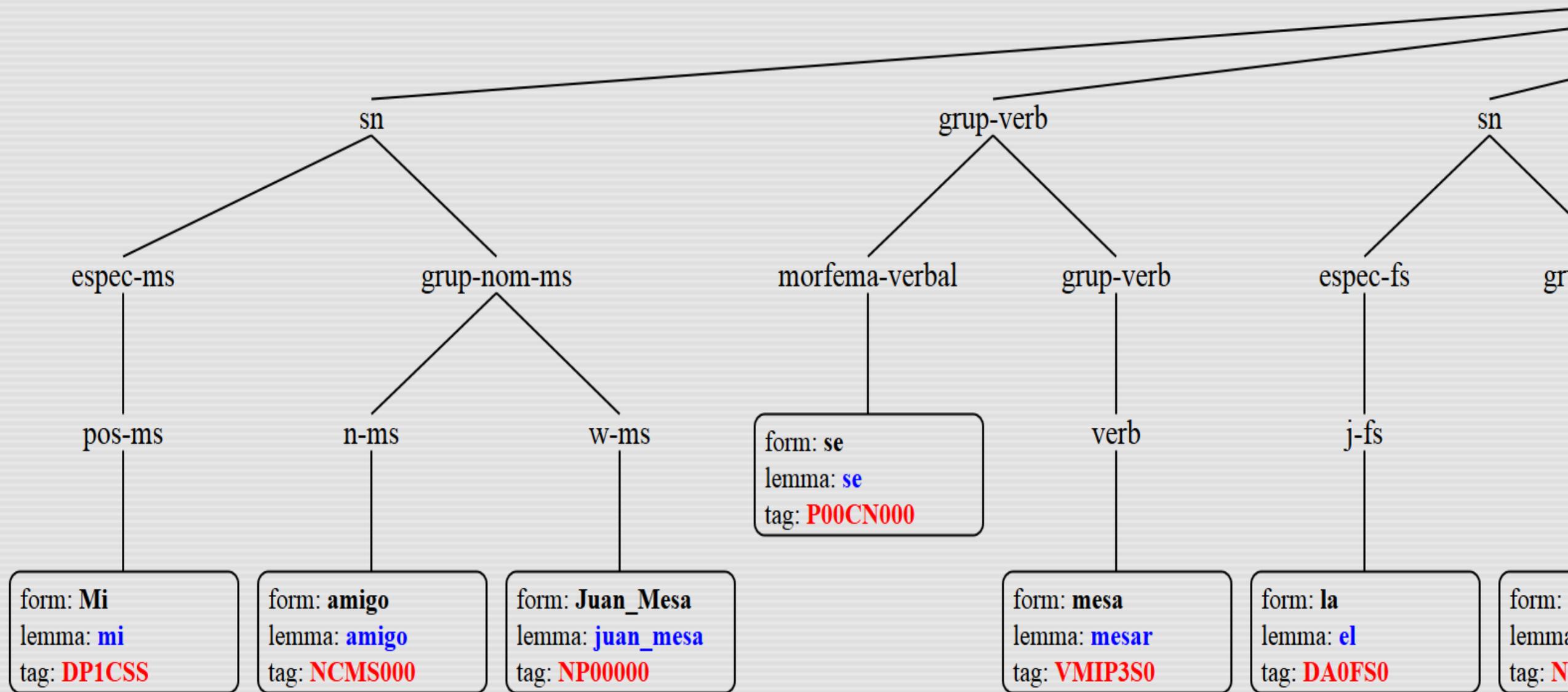
Análisis sintáctico (Parsing)

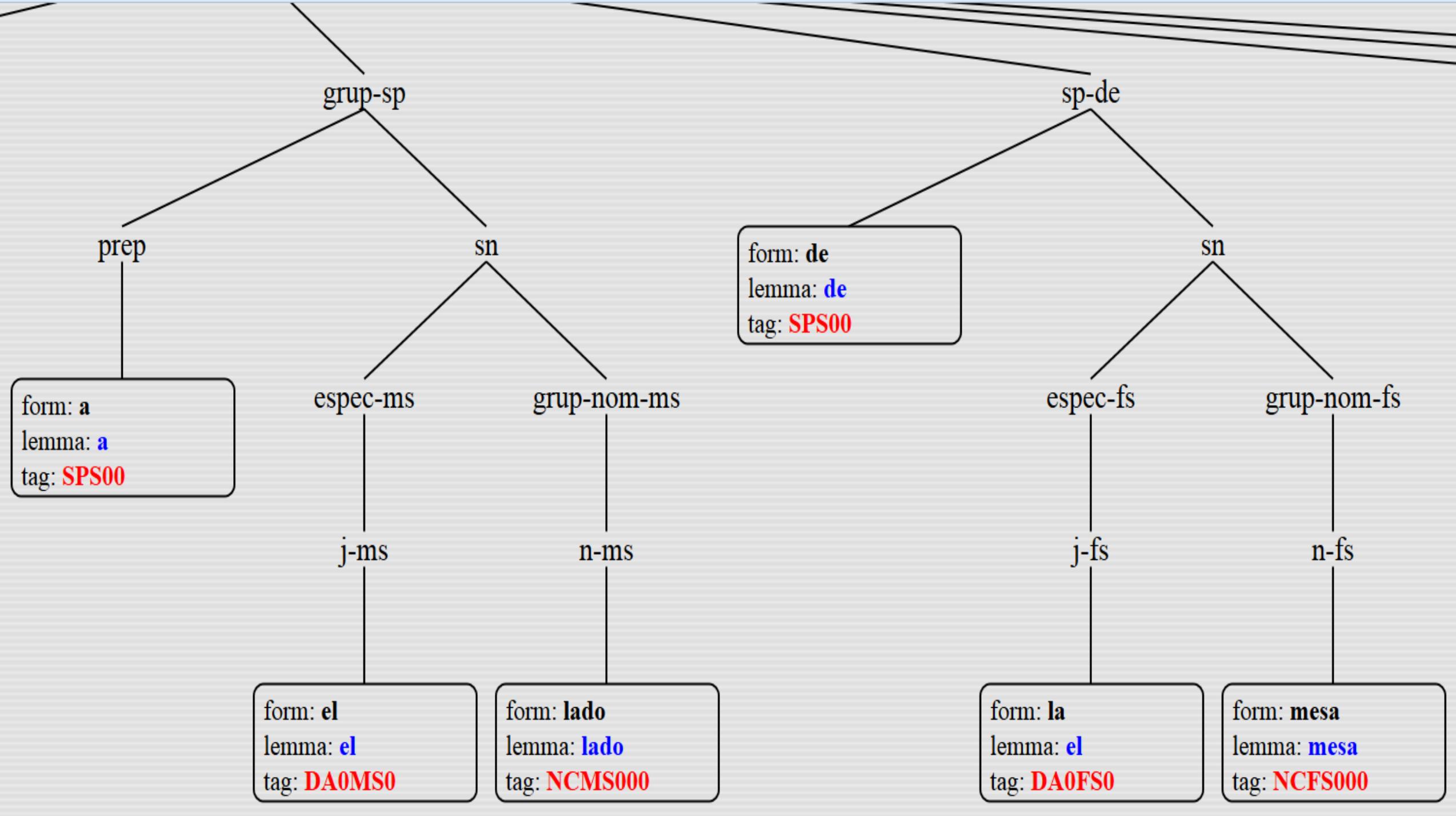
- Análisis sintáctico estructural (árbol).

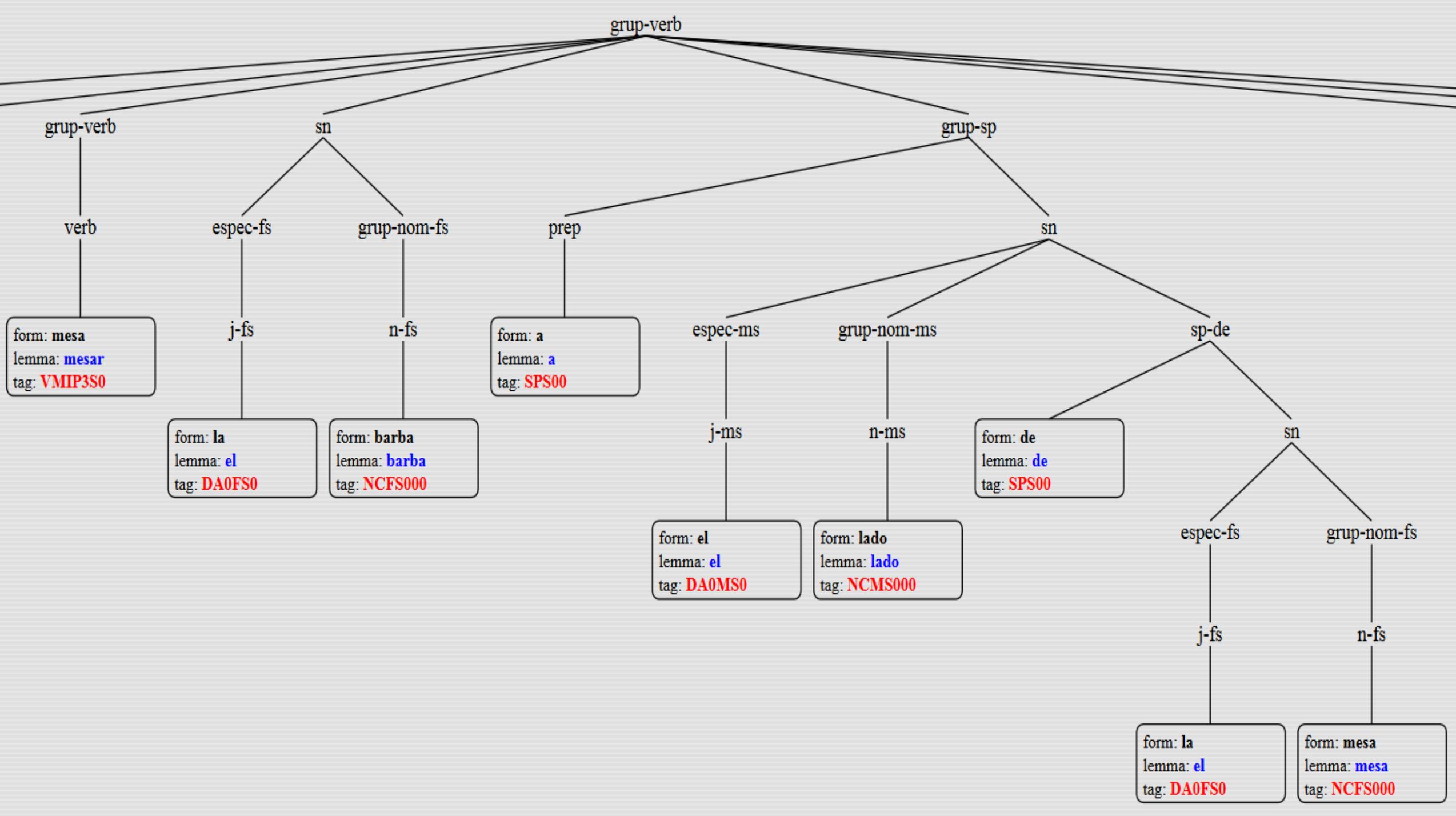
Análisis sintáctico de dependencias (Dependency parsing)

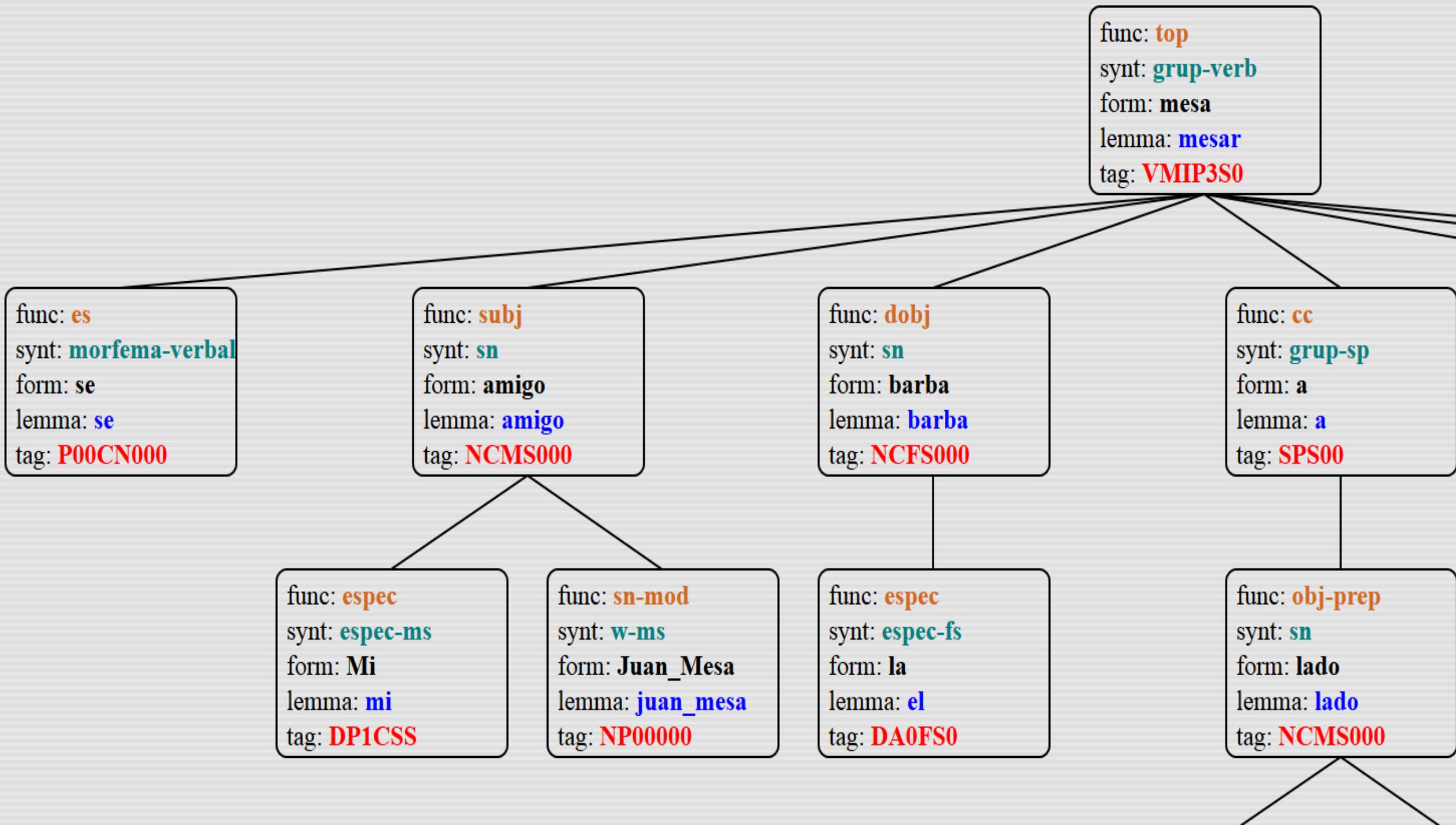
- Análisis de funciones sintácticas (árbol de dependencias).

Sentence #1









Fin