

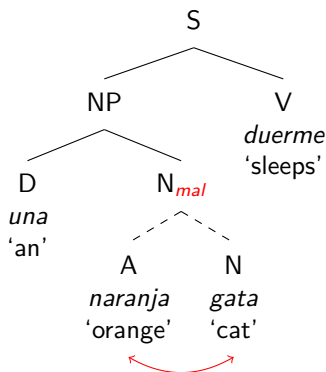
Grammar checking with SRG

for the DELPH-IN summit

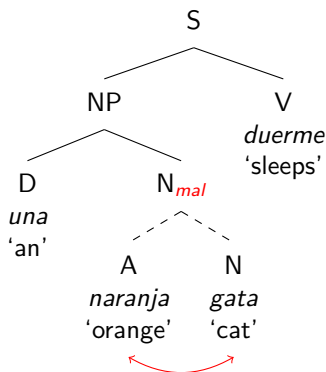
Olga Zamaraeva, Carlos Gómez-Rodríguez, Margarita Alonso-Ramos
Department of Informatics/CITIC, Department of Philology
Universidade da Coruña

July 17 2022

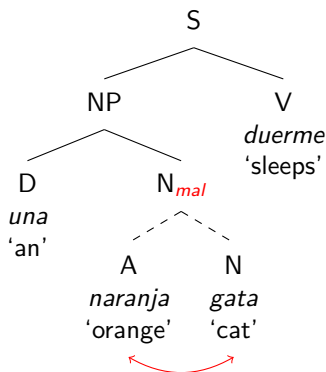
New grant for DELPH-IN work



New grant for DELPH-IN work

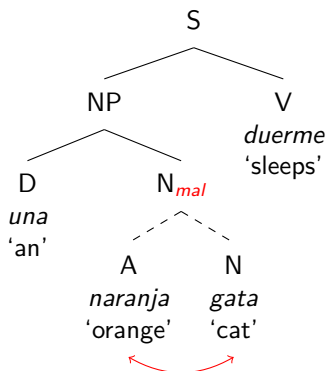


New grant for DELPH-IN work



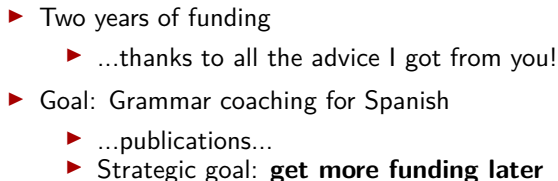
New grant for DELPH-IN work

Grammar checking
with SRG
for the DELPH-IN
summit

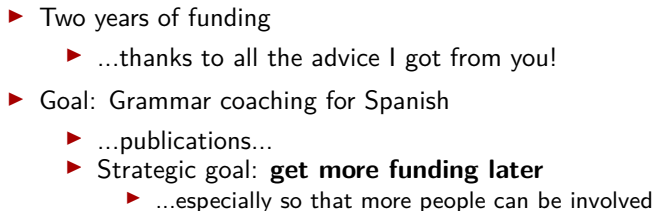


- ▶ Two years of funding
 - ▶ ...thanks to all the advice I got from you!
- ▶ Goal: Grammar coaching for Spanish
 - ▶ ...publications...

Grammar checking
with SRG
for the DELPH-IN
summit



Grammar checking
with SRG
for the DELPH-IN
summit



Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules

Determining the scope: Types of tasks/expertise

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules
- ▶ Designing feedback

Determining the scope: Types of tasks/expertise

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules
- ▶ Designing feedback
 - ▶ L2 publication

Determining the scope: Types of tasks/expertise

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules
- ▶ Designing feedback
 - ▶ L2 publication
 - ▶ Related work at UDC but not for syntax

Determining the scope: Types of tasks/expertise

Grammar checking
with SRG
for the DELPH-IN
summit

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules
- ▶ Designing feedback
 - ▶ L2 publication
 - ▶ Related work at UDC but not for syntax
- ▶ Developing and serving the application

Determining the scope: Types of tasks/expertise

- ▶ Developing the SRG for coverage
 - ▶ ...probably the only area where I have expertise*
 - ▶ *...not that I am Spanish speaker
- ▶ Designing mal-rules
 - ▶ L2 publication venues
 - ▶ Alonso-Ramos works with learner corpora but focuses on collocations (hard to model syntactically?)
 - ▶ Work st UC Davis on a learner corpus but for ML
- ▶ Integrating mal-rules
- ▶ Designing feedback
 - ▶ L2 publication
 - ▶ Related work at UDC but not for syntax
- ▶ Developing and serving the application
- ▶ Testing and supporting the application

- ▶ SRG
 - ▶ Coverage: 35.5% over the Ancora corpus (2010)
- ▶ Annotated learner corpora:
 - ▶ Davidson et al. (US), 33,950 sentences
 - ▶ <https://github.com/ucdaviscl/cowsl2h>
 - ▶ Yo vivo en [el]{la}<ga:fm:art>ciudad.
 - ▶ CORELE (Spain)
 - ▶ Oral
 - ▶ Only available through web interface?
 - ▶ Only 10 participants
 - ▶ But possible to get examples of many error categories
 - ▶ Aprescrllov (Spain), 1 mln words
 - ▶ Annotated for L1
 - ▶ Presumably for some other stuff but hard to say

- ▶ Gender and number agreement
- ▶ the use of *a* with direct objects
- ▶ Ser/estar??..
 - ▶ not sure if modelable syntactically
 - ▶ but there is some stuff in SRG already
 - ▶ may be interesting in terms of RQ
- ▶ Adpositions?..
- ▶ Other things:
<http://cartago.llf.uam.es/exist/rest//db/corpus/>