

On the problems of creating a golden standard of inflected forms in Portuguese

Morfolimpíadas

- the first evaluation contest for Portuguese
- evaluate state-of-the-art morphological analyzers
- measure progress in Portuguese language engineering
- define boundaries of Portuguese morphology
- start a new collaborative paradigm

What is the golden list?

- an evaluation resource to be used as a standard to rank morphological analyzers according to precision and coverage
- a set of correct analyses of 655 Portuguese inflected forms
- a means to build theoretical and practical consensus to evaluate morphological analyzers

Compilation principles

- must be fair
- only common dimensions can be compared
- must allow flexible metrics
- must reflect interesting cases
- construction based on consensus

What makes it difficult to compare?

- number of analyses (I vs. S and P)
- different tokenization
- capitalization and punctuation differently treated
- derivation or not? How to encode it?
- PoS disagreement
- theoretical disagreement on the meaning of a label (past participle)

Golden List Examples

Brava:ADJ+bravo+.+S+.+F+.+.

×

Chaves:PROP+Chaves+.+S+.+.

Chaves:SUB+chaves+.+P+.+F+.+.

×

Marco:PROP+Marco+.+S+.+M+.+.

Marco:SUB+marco+.+S+.+M+.+.

Marco:V+marcar+PR_I+S+1+.+.

×

Marinho:ADJ+marinho+.+S+.+M+.+.

Marinho:PROP+Mário+.+S+.+M+.+DIM+

Marinho:SUB+marinho+.+S+.+M+.+.

×

Miss:SUB+miss+.+S+.+F+.+.

×

Papa:PROP+Papa+.+S+.+M+.+.

Papa:SUB+papa+.+S+.+F+.+.

Papa:SUB+papa+.+S+.+M+.+.

Papa:V+papar+IMP+S+2+.+.

Papa:V+papar+PR_I+S+3+.+.

×

antimotim:ADJ+antimotim+.+P+.+I+.+.

antimotim:ADJ+antimotim+.+S+.+I+.+.

×

celular:ADJ+celular+.+S+.+I+.+.

celular:SUB+celular+.+S+.+M+.+.

×

expôr+X+?+.+.

×

frete:SUB+frete+.+S+.+M+.+.

frete:V+fretar+PR_C+S+1+.+.

frete:V+fretar+PR_C+S+3+.+.

×

inalterada:ADJ+inalterado+.+S+.+F+.+.

inalterada:ADJ+alterado+.+S+.+F+.+.

inalterada:V+alterar+PP+S+.+F+.+.

×

retratado:ADJ+retratado+.+S+.+M+.+.

retratado:ADJ+tratado+.+S+.+M+.+.

retratado:V+retratar+PP+S+.+M+.+.

retratado:V+tratar+PP+S+.+M+.+.

×

utilizador:ADJ+utilizador+.+S+.+M+.+.

utilizador:ADJ+utilizar+.+S+.+M+.+.

utilizador:SUB+utilizador+.+S+.+M+.+.

utilizador:SUB+utilizar+.+S+.+M+.+.

gender of proper names

relative frequency of analyses

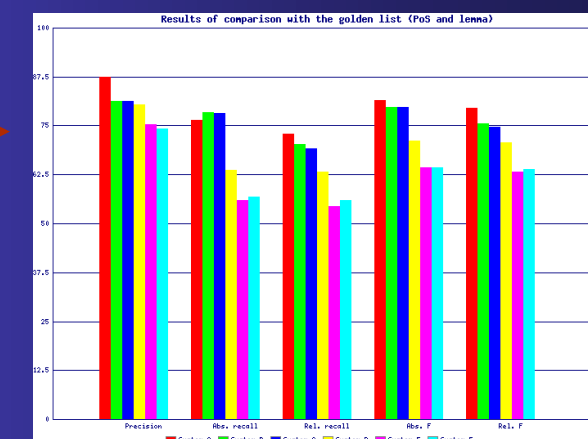
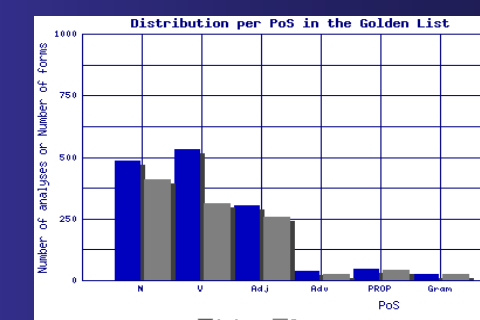
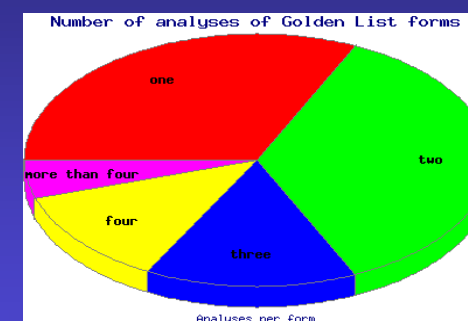
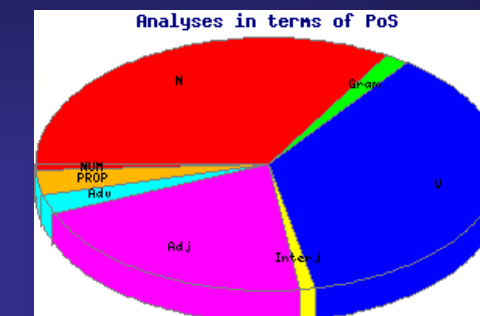
one or two analyses?

Several comparisons performed

- PoS and lemma
- everything but lemma
- only PoS

variant

what is the right lemma?



Conclusions

- there is still a lot of disagreement on both theoretical and practical issues dealing with the behavior of a Portuguese morphological analyzer, but we achieved a first "map"
- we created a public resource, which allows others to compare and define their own evaluation metrics and standards, to be used in next evaluation contests, and to try their impact on systems ranking