

# Efficiency, Robustness, Precision: Balance and control

**Dan Flickinger**

CSLI, Stanford University

DELPH-IN Summit 2019

Cambridge

17 July 2019

# Efficiency

- Grammarians

  - Removing unwanted ambiguity in chart (spurious phrases)

  - Reducing FS size by adding underspecified types for feats

- Developers

  - Using ubertagging to remove unlikely lexical edges from chart

  - Using statistical models to remove phrasal edges from chart

  - Making engines more efficient (e.g. generalizing types in unif)



# Robustness

- Grammarians
  - POS-tag-based unknown word additions
  - Bridging rules
  - Peninsula rules (NP, VP)
- Developers: PCFG trained on parsed corpora
  - Either add best PCFG analysis edges to normal chart
  - Or back off to PCFG when normal grammar fails to parse



# Precision

- For comparison with treebanked dependencies, need to manufacture gold MRSs for unparsable sentences in a profile

Slightly rephrase problematic sentences to make them parsable

Insert phrasal boundaries to overcome efficiency bounds



# Alchemy for evaluation profiles

1000007800300@unknown@formal@none@1@S@Ladino, which is essentially medieval Spanish and closer to modern Spanish than any other language, is spoken by many descendants of the Sephardi Jews who were expelled from Spain in the 15th century.@@@1@33@@oe@26-8-2013

1000007800300@unknown@formal@none@1@S@Ladino, which is essentially medieval Spanish and is closer to modern Spanish than any other language, is spoken by many descendants of the Sephardi Jews who were expelled from Spain in the 15th century.@@@1@33@@oe@26-8-2013



# Alchemy for evaluation profiles

1000007800440@unknown@formal@none@1@S@ Mutual intelligibility  
between Spanish and French or Romanian is even lower (lexical  
similarity being respectively 75% and 71%): comprehension of  
Spanish by French speakers who have not studied the language is as  
low as an estimated 45% - the same as of English.@@@1@43@@oe@26-8-2013

1000007800440@unknown@formal@none@1@S@[ ( Mutual intelligibility  
between Spanish and French or Romanian [ ) ] is even lower (lexical  
similarity being respectively 75% and 71%): [ ( comprehension of  
Spanish by French speakers who have not studied the language is as  
low as an estimated 45% - the same as of English. [ ) ]@@@1@43@@oe@26-8-



# User controls

```
ace -g ~/erg/erg.dat -n 1  
--pcfg=/home/danf/erg/etc/all-treebanks-gp0-2018.pcfg  
--max-chart-megabytes=9000 --max-unpack-megabytes=10000  
--timeout=100 --max-words=150 --ubertagging=0.00001
```



# User controls: Robustness

```
ace -g ~/erg/erg.dat -n 1  
--pcfg=/home/danf/erg/etc/all-treebanks-gp0-2018.pcfg  
--max-chart-megabytes=9000 --max-unpack-megabytes=10000  
--timeout=100 --max-words=150 --ubertagging=0.00001
```





# User controls: Efficiency

```
ace -g ~/erg/erg.dat -n 1  
--pcfg=/home/danf/erg/etc/all-treebanks-gp0-2018.pcfg  
--max-chart-megabytes=9000 --max-unpack-megabytes=10000  
--timeout=100 --max-words=150 --ubertagging=0.00001
```

