POS Tagging

• POS Tagging
• Rule-based taggers
• Statistical taggers
• Hybrid approaches
Words taken isolatedly are ambiguous regarding its POS

Yo  bajo  con  el  hombre bajo a

PP  VM  SP  TD  NC  VM  NC
VM  VM  AQ  SP
AQ  NC  SP

tocar el  bajo  bajo  la escalera .

VM  TD  VM  TD  NC  FP
VM  VM  AQ  NC
AQ  NC  PP
NC  SP  SP
Most of words have a unique POS within a context

<table>
<thead>
<tr>
<th>Yo</th>
<th>bajo</th>
<th>con</th>
<th>el</th>
<th>hombre bajo a</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>VM</td>
<td>SP</td>
<td>TD</td>
<td>NC</td>
</tr>
<tr>
<td>VM</td>
<td>AQ</td>
<td>NC</td>
<td>SP</td>
<td>VM</td>
</tr>
<tr>
<td>NC</td>
<td>AQ</td>
<td>SP</td>
<td></td>
<td>NC</td>
</tr>
<tr>
<td>SP</td>
<td></td>
<td></td>
<td></td>
<td>SP</td>
</tr>
</tbody>
</table>

tocar el  bajo  bajo  la escalera .

| VM   | TD    | VM    | VM    | TD    | NC    | FP |
| VM   |       | AQ    | AQ    |       | NC    | PP |
| AQ   |       | NC    | NC    |       | SP    |   |
The goal of a POS tagger is to assign each word the most likely within a context.

- Rule-based
- Statistical
- Hybrid
POS Tagging

\[ W = w_1 \ w_2 \ldots \ w_n \text{ sequence of words} \]
\[ T = t_1 \ t_2 \ldots t_n \text{ sequence of POS tags} \]

For each word \( w_i \), only some of the tags can be assigned (except the unknown words). We can get them from a lexicon or a morphological analyzer.

Tagset.

Open vs closed categories

\[ f : W \rightarrow T = f(W) \]
• Knowledge-driven taggers
• Usually rules built manually
• Limited amount of rules ($\approx 1000$)
• LM and smoothing explicitly defined.
Rule-based taggers

+ Linguistically motivated rules
+ High precision
  + ej. EngCG 99.5%
- High development cost
- Not transportable
- Time cost of tagging

- TAGGIT, Green, Rubin, 1971
- TOSCA, Oosdijk, 1991
- AMBILIC, de Yzaguirre et al, 2000
A CG consists of a sequence of subgrammars each one consisting of a set of restrictions (constraints) which set context conditions

- ej. (@w =0 VFIN (-1 TO))
  - Discards POS VFIN when the previous word is “to”

**ENGCG**

- ENGTWOL
- Reductionist POS tagging
  - 1,100 constraints
  - 93-97% of the words are correctly disambiguated
  - 99.7% accuracy
  - Heuristic rules can be applied over the rest
    - 2-3% residual ambiguity with 99.6% precision

- CG syntactic
• LM and smoothing automatically learned from tagged corpora (supervised learning).
• Data-driven taggers
• Statistical inference
• Techniques coming from speech processing
POS Tagging

Statistical POS taggers

- Well founded theoretical framework
- Simple models.
- Acceptable precision
  - > 97%
- Language independent
  - Learning the model
    - Sparseness
  - Less precision

- CLAWS, Garside et al, 1987
- De Rose, 1988
- Church, 1988
- Cutting et al, 1992
- Merialdo, 1994
Statistical POS taggers

- N-gram
  - smoothing
  - interpolation

- HMM

- ML
  - Supervised learning
    - MLE
  - Semi-supervised
    - Forward-Backward, Baum-Welch (EM Expectation Maximization)

Charniak, 1993
Jelinek, 1998
Manning, Schütze, 1999
3-gram tagger

\[ \arg \max_{t_1 \ldots t_n} P(t_1, \ldots, t_n \mid w_1, \ldots, w_n) \approx \prod_{k=1}^{n} P(t_k \mid t_{k-2}, t_{k-1}) \cdot P(w_k \mid t_k) \]

- Contextual probability (trigrams)
- Lexical Probability
• Hidden States associated to n-grams
• Transition probabilities restricted to valid transitions:
  • *BC -> BC*
• Emision probabilities restricted by lexicons
• Transformation-based, error-driven
  • Based on rules automatically acquired

• Maximum Entropy
  • Combination of several knowledge sources
  • No independence is assumed
  • A high number of parameters is allowed (e.g. lexical features)
• Based on transformation rules that correct errors produced by an initial HMM tagger

• rule
  • change label A into label B when ...
  • Each rule corresponds to the instantiation of a template

• templates
  • The previous (following) word is tagged with Z
  • One of the two previous (following) words is tagged with Z
  • The previous word is tagged with Z and the following with W
  • ...

• Learning of the variables A,B,Z,W through an iterative process that chooses at each iteration the rule (the instantiation) correcting more errors.
• Decision trees
  • Supervised learning
  • ej. TreeTagger
• Case-based, Memory-based Learning
  • IGTREE
• Relaxation labelling
  • Statistical and linguistic constraints
  • ej. RELAX

Other complex systems

Black, Magerman, 1992
Magerman 1996
Màrquez, 1999
Màrquez, Rodríguez, 1997

TiMLB
Daelemans et al, 1996

Padrò, 1997
Márquez’s Tree-tagger

IN/RB ambiguity
IN (preposition)
RB (adverb)

statistical interpretation:

\[ \hat{P}( \text{RB} | \text{word} = “A/as” \& \text{tag}(+1) = \text{RB} \& \text{tag}(+2) = \text{IN}) = 0.987 \]

\[ \hat{P}( \text{IN} | \text{word} = “A/as” \& \text{tag}(+1) = \text{RB} \& \text{tag}(+2) = \text{IN}) = 0.013 \]
Combining taggers

- Combination of LM in a tagger
  - STT+
  - RELAX
- Combination of taggers through votation
  - bootstrapping
- Combinación de classifiers
  - bagging (Breiman, 1996)
  - boosting (Freund, Schapire, 1996)

References:
- Màrquez, Rodríguez, 1998
- Màrquez, 1999
- Padrò, 1997
- Màrquez et al, 1998
- Brill, Wu, 1998
- Màrquez et al, 1999
- Abney et al, 1999
POS Tagging

Màrquez’s STT+

Language Model

Lexical probs.  +  Contextual probs.  +  N-grams

Viterbi algorithm

Disambiguation

Raw text → Morphological analysis → Viterbi algorithm → Tagged text

Morphological analysis → Viterbi algorithm → Tagged text
POS Tagging

Language Model

N-grams + Set of constraints + Linguistic rules

Relaxation Labelling (Padró, 1996)

Disambiguation

Raw text → Morphological analysis → Relaxation Labelling (Padró, 1996) → Tagged text

Set of constraints

Morphological analysis

N-grams

Linguistic rules

Padró’s Relax

POS Tagging