POS Tagging 1

- POS Tagging
- Rule-based taggers
- Statistical taggers
- Hybrid approaches

POS Tagging 2

Words taken isolatedly are ambiguous regarding its POS

Yo	bajo	con	el	hombre	bajo	a
PP	VM	SP	TD	NC	VM	NC
	VM				VM	SP
	AQ				AQ	
	NC				NC	
	SP				SP	

tocar	el	bajo	bajo	la	escalera	.
VM	TD	VM	VM	TD	NC	FP
VM		VM	VM	NC		
		AQ	AQ	PP		
		NC	NC			
		SP	SP			

POS Tagging 3

Most of words have a unique POS within a context

Yo	bajo	con	el	hombre	bajo	a
PP	**VM**	SP	TD	NC	VM	NC
	VM				VM	SP
	AQ				**AQ**	
	NC				NC	
	SP				SP	

tocar	el	bajo	bajo	la	escalera	.
VM	TD	VM	VM	**TD**	NC	FP
VM		VM	VM	NC		
		AQ	AQ	PP		
		NC	NC			
		SP	**SP**			

POS Tagging 4

Pos taggers

The goal of a POS tagger is to assign each word the most likely within a context

- Rule-based
- Statistical
- Hybrid

POS Tagging 5

$W \quad = w_1\ w_2\ \dots\ w_n$ sequence of words
$T \quad = t_1\ t_2\ \dots t_n$ 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

$$\boxed{f : W \rightarrow T = f(W)}$$

Rule-based taggers

- Knowledge-driven taggers
- Usually rules built manually
- Limited amount of rules (≈ 1000)
- LM and smoothing explicitely defined.

POS Tagging 7

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
- Constraint Grammars, EngCG, Voutilainen,1994, Karlsson et al, 1995
- AMBILIC, de Yzaguirre et al, 2000

POS Tagging 8

Constraint Grammars CG

- 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 correctily disambiguated
    - 99.7% accuracy
    - Heuristic rules can be applied over the rest
      - 2-3% residual ambiguity with 99.6% precision
  - CG syntactic

POS Tagging 8

Statistical POS taggers

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

POS Tagging 9

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

POS Tagging 10

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

POS Tagging 11

3-gram tagger

$$\arg\max_{t_1 \cdots t_n} P(t_1, \ldots, t_n \mid w_1, \ldots, w_n) \approx$$

$$\prod_{k=1}^{n} \underbrace{P(t_k \mid t_{k-2}, t_{k-1})}_{\text{Contextual probability (trigrams)}} \cdot \underbrace{P(w_k \mid t_k)}_{\text{Lexical Probability}}$$

POS Tagging 12

HMM tagger

- Hidden States associated to n-grams
- Transition probabilities restricted to valid transitions:
  - *BC -> BC*
- Emision probabilities restricted by lexicons

POS Tagging 13

Hybrid systems

- Transformation-based, error-driven
  - Based on rules automatically acquired

Brill, 1995
Roche,Schabes, 1995

- Maximum Entropy
  - Combination of several knowledge sources
  - No independence is assumed
  - A high number of parameters is allowed (e.g. lexical features)

Ratnaparkhi, 1998,
Rosenfeld, 1994
Ristad, 1997

POS Tagging 14

Brill’s system

- Based on transformation rules that corerect errors produced by an initial HMM tagger
- rule
  - change label A into label B when ...
  - Each rule corresponds to the instantiation of a templete
- templetes
  - 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 choose at each iteration the rule (the instanciation) correcting more errors.

Other complex systems

- Decision trees
  - Supervised learning
  - ej. TreeTagger
- Case-based, Memory-based Learning
  - IGTree
- Relaxation labelling
  - Statistical and linguistic constraints ej. RELAX

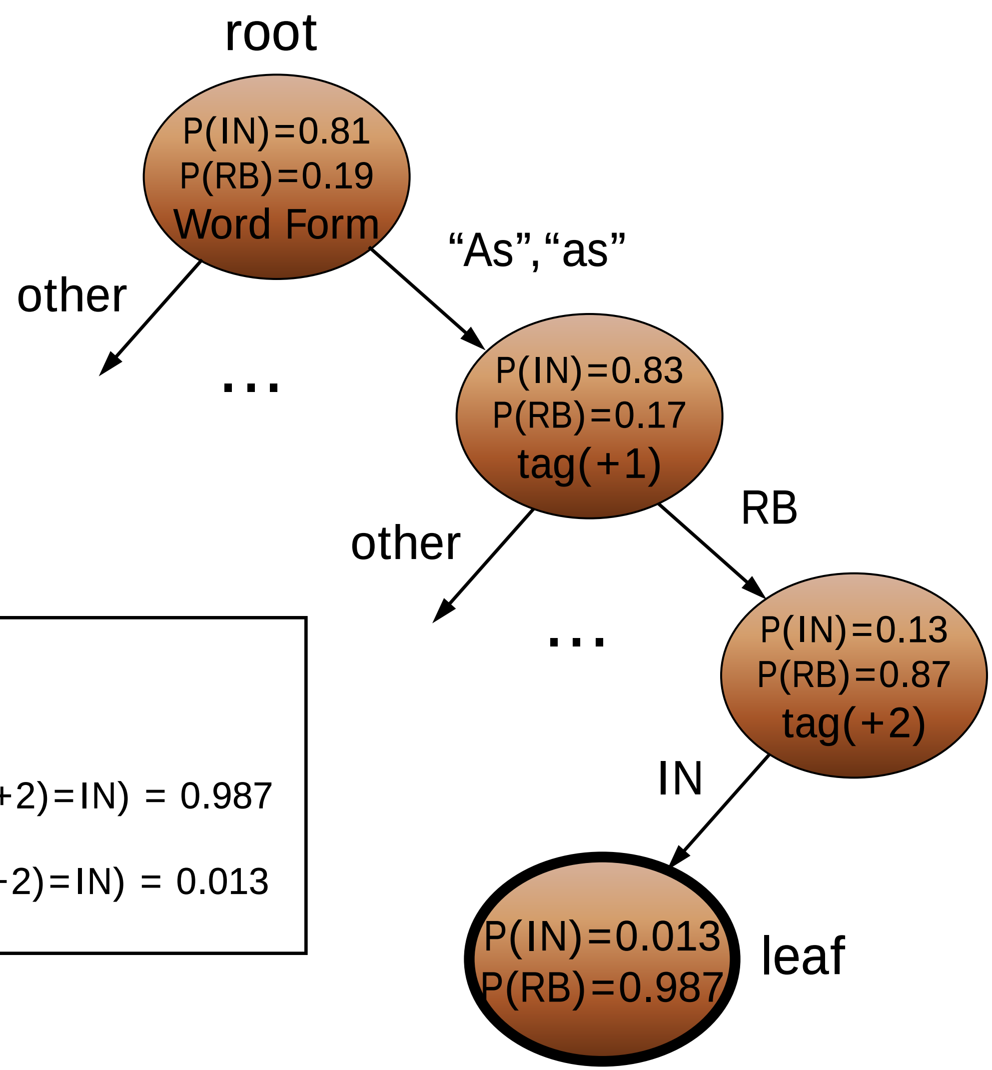
Black,Magerman, 1992
Magerman 1996
Màrquez, 1999
Màrquez, Rodríguez, 1997

TiMBL
Daelemans et al, 1996

Padrò, 1997

Màrquez's Tree-tagger

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



statístical interpretation :

$\hat{P}$(RB | word="A/as" & tag(+1)=RB & tag(+2)=IN) = 0.987

$\hat{P}$(IN | word="A/as" & tag(+1)=RB & tag(+2)=IN) = 0.013

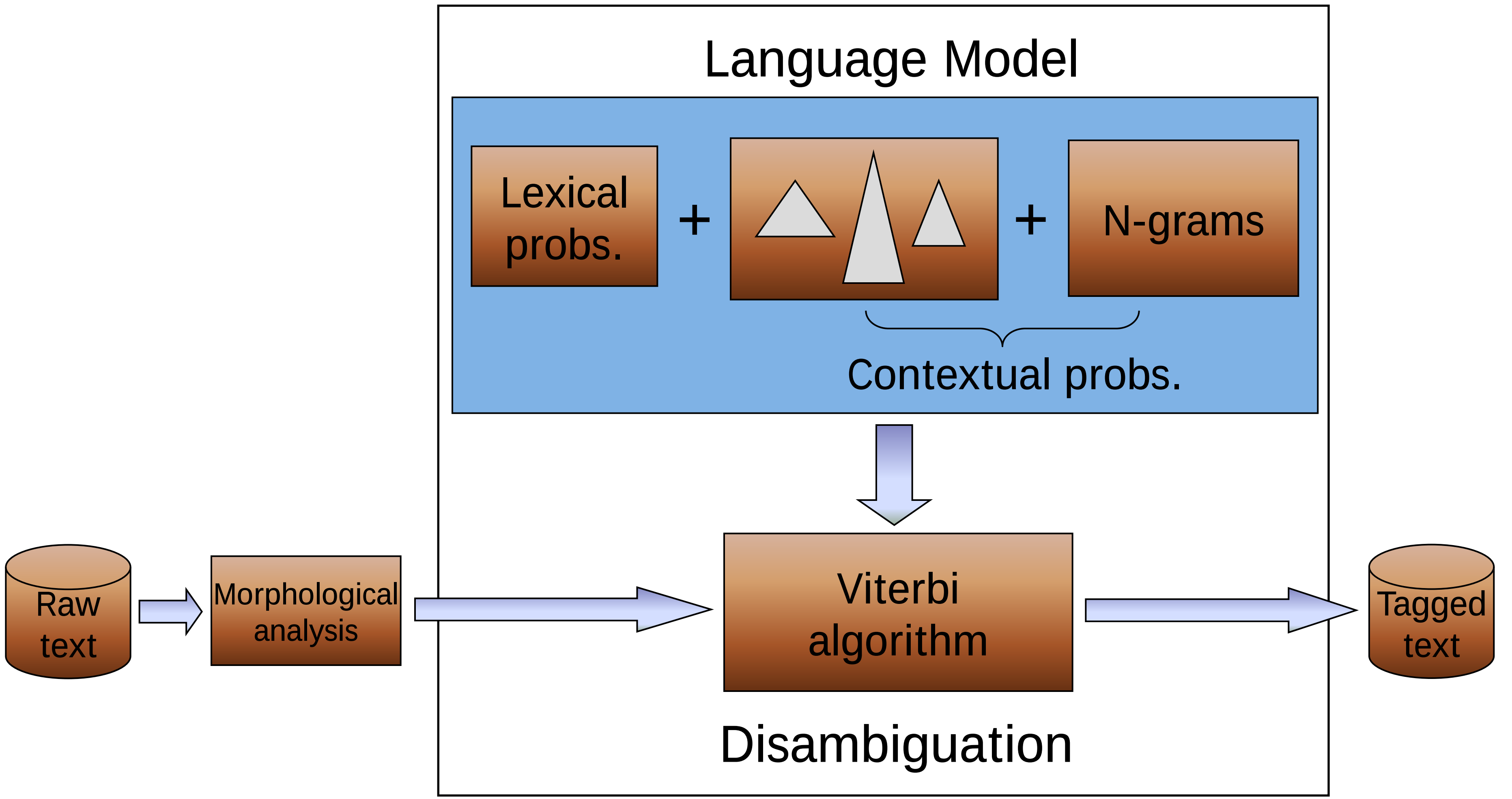
POS Tagging 17

Combining taggers

- Combination of LM in a tagger → Màrquez, Rodríguez, 1998; Màrquez, 1999; Padrò, 1997
  - STT+
  - RELAX
- Combination of taggers through votation → Màrquez et al, 1998
  - bootstrapping
- Combinación of classifiers
  - bagging (Breiman, 1996)
  - boosting (Freund, Schapire, 1996) → Brill, Wu, 1998; Màrquez et al, 1999; Abney et al, 1999

POS Tagging 18

Màrquez's STT+



POS Tagging 19

Padró's Relax