ILLINOIS INSTITUTE OF TECHNOLOGY

Transformation-Based Error-Driven Learning and Natural Language Processing : A Case Study in Part-of-Speech Tagging

Eric Brill (1995)

Jiranun Jiratrakanvong

A20337992

Agenda

ILLINOIS INSTITUTE OF TECHNOLOGY

- What is "**Transformation**"?
- Transformation-Based Error-Driven Learning
- A Comparison With Decision Trees
- Part of Speech Tagging : A Case Study in Transformation-Based Error-Driven Learning

What is "**Transformation**" ?

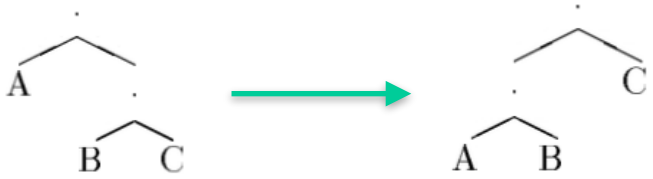
ILLINOIS INSTITUTE OF TECHNOLOGY

» **Triggering Environment**
e.g. The preceding word is a determiner

» **Rewrite Rule**
e.g. Change the tag from MODAL to NOUN

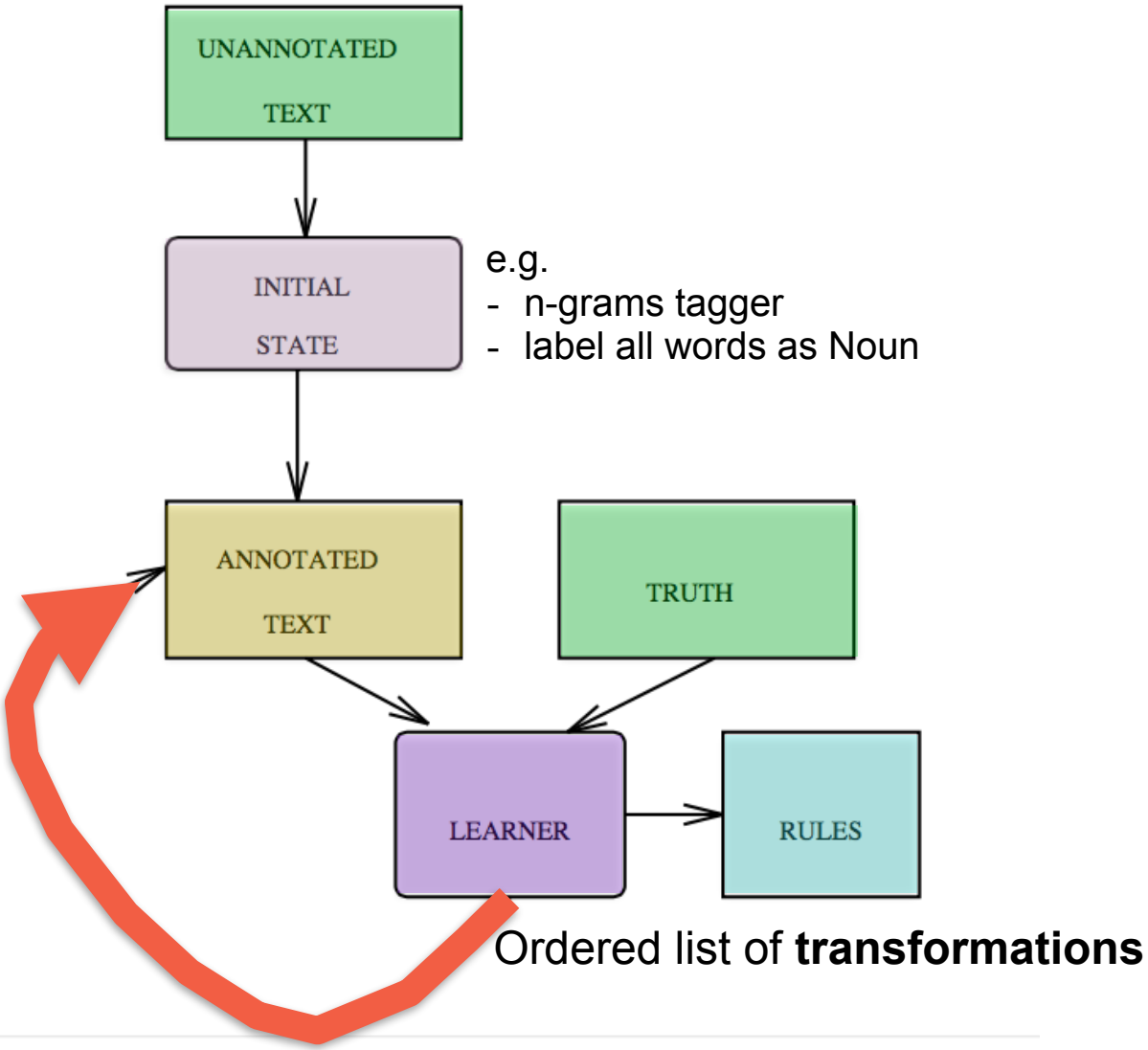
» <u>For example</u> : "*The can rusted.*"

- *Before* : *The/DT can/**MD** rusted/VB ./.*
- *After* : *The/DT can/**NN** rusted/VB ./.*



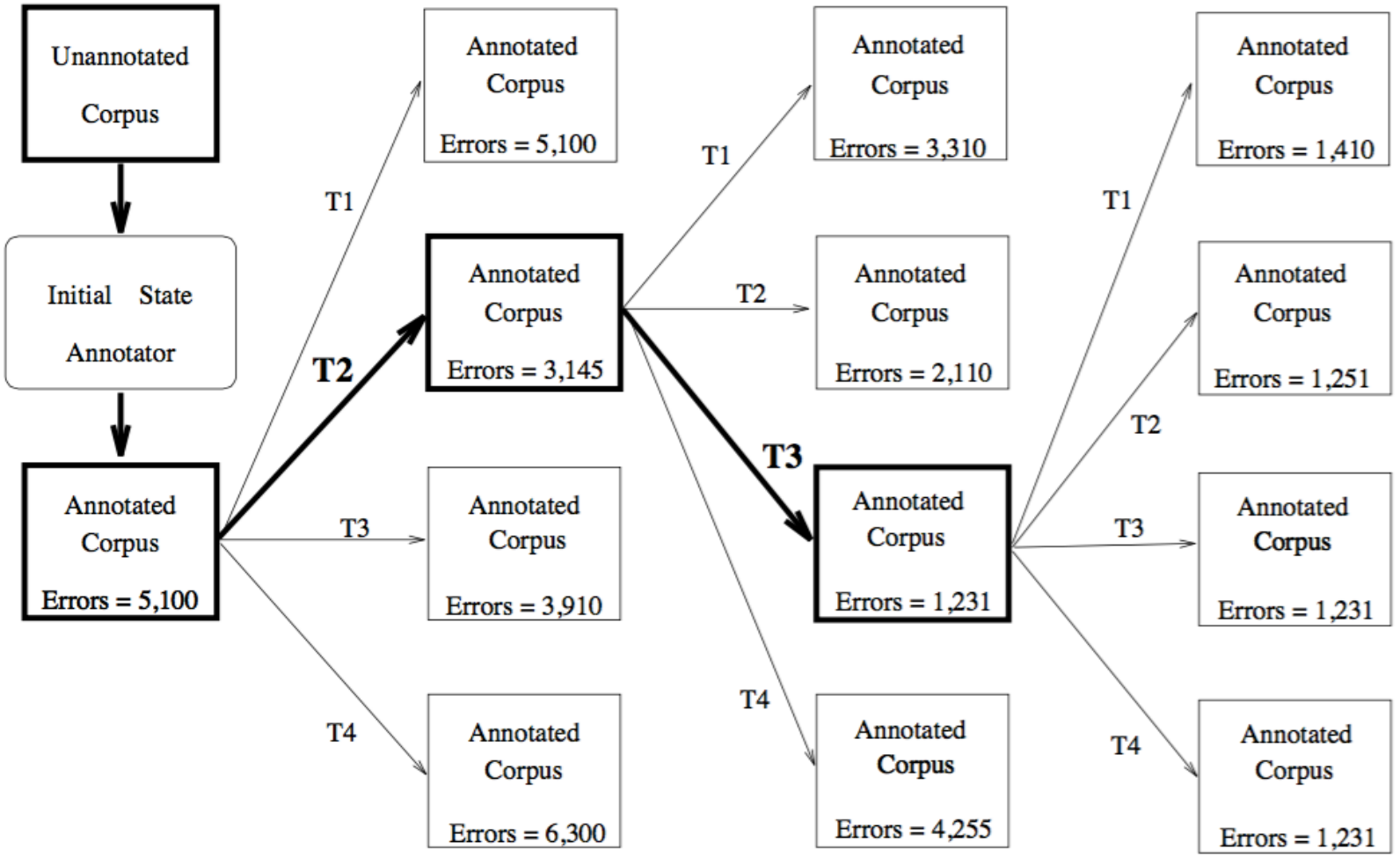
Transformation-Based Error-Driven Learning

ILLINOIS INSTITUTE OF TECHNOLOGY



Transformation-Based Error-Driven Learning

ILLINOIS INSTITUTE OF TECHNOLOGY



An Example of Transformation-Based Error-Driven Learning

Transformation-Based Error-Driven Learning

ILLINOIS INSTITUTE OF TECHNOLOGY

We must specify

- **The initial state-annotator**
- **The space of allowable transformations**
  (rewrite rules and triggering environments)
- **The objective function**
  for comparing the corpus to the truth and choosing transformation
- **Two additional parameters**
  about how each transformation apply

Transformation-Based Error-Driven Learning

ILLINOIS INSTITUTE OF TECHNOLOGY

TWO ADDITIONAL PARAMETERS

1. What's the **order** in which transformations are applied to a corpus. R → L or L → R

2. Whether the effect of a transformation is applied **immediately** or **after** the corpus has been triggered

Transformation-Based Error-Driven Learning

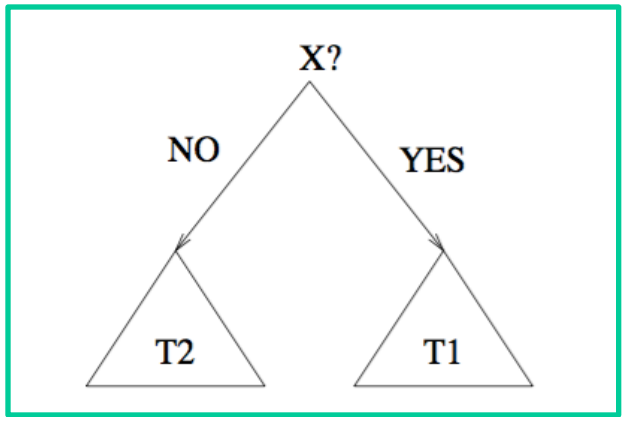
ILLINOIS INSTITUTE OF TECHNOLOGY

» For Example
  – **sequence** : A A A A A A
  – **transformation** : *Change from A to B if the preceding label is A*
  – **results** can be
    • A B B B B B
    • A B A B A B

A Comparison With Decision Trees

ILLINOIS INSTITUTE OF TECHNOLOGY

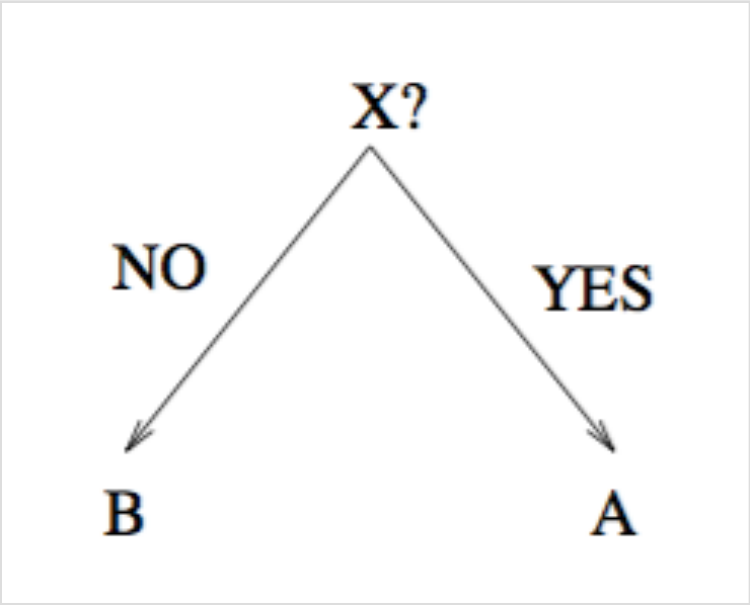
- Decision Trees ⊆ Transformation Lists
- Decision Trees ≠ Transformation Lists
- Some Practical Differences Between Decision Trees and Transformation Lists



Transformation:
Trigger Environments
…
Rewrite rules
…

Decision Trees ⊆ Transformation Lists

ILLINOIS INSTITUTE OF TECHNOLOGY



1. Label with S
2. If X, then S → A
3. S → B

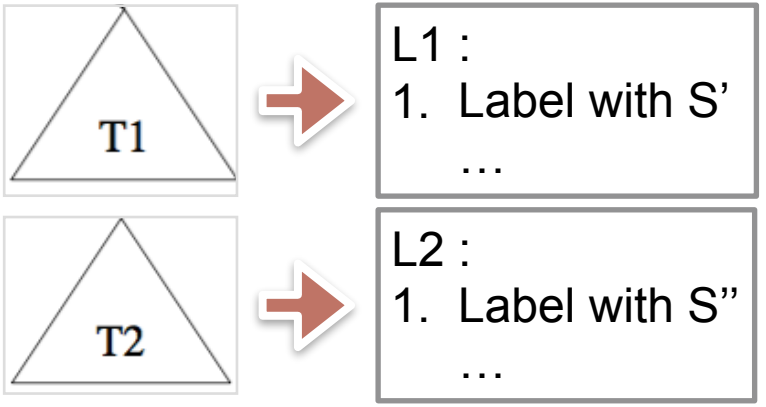
Decision tree can be converted into a transformation list

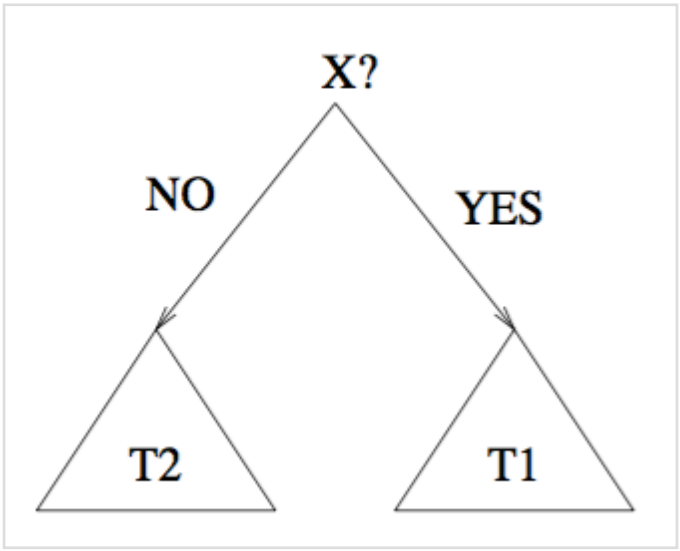
Decision Trees ⊆ Transformation Lists

ILLINOIS INSTITUTE OF TECHNOLOGY

Assume :

- T_n has corresponding transformation lists L_n
- Labels names in L_1 and L_2 are not the same





T3

L3 :
1. Label with S
2. If X, then S → S'
3. S → S"

Decision Trees ≠ Transformation Lists

ILLINOIS INSTITUTE OF TECHNOLOGY

» **Given** : A A A A A A A A A A

» **Problem** : Classify a character based on whether the position index of a character is divisible by 4

» **Expected Result** :

A/yes A/no A/no A/no A/yes A/no A/no A/no A/yes A/no

0 1 2 3 4 5 6 7 8 9

Decision Trees ≠ Transformation Lists

ILLINOIS INSTITUTE OF TECHNOLOGY

1. Label with S
   A/**S** A/**S** A/**S** A/**S** A/**S** A/**S** A/**S** A/**S** A/**S** A/**S**
2. No previous character, S → F
   A/**F** A/S A/S A/S A/S A/S A/S A/S A/S A/S
3. Two to the left is F, S → F
   A/F A/S A/**F** A/S A/**F** A/S A/**F** A/S A/**F** A/S
4. Two to the left is F, F → S
   A/F A/S A/**S** A/S A/F A/S A/**S** A/S A/F A/S
5. F → yes
6. S → no

A/yes A/no A/no A/no A/yes A/no A/no A/no A/yes A/no

Some Practical Differences

ILLINOIS INSTITUTE OF TECHNOLOGY

- Data problems
- Immediate Classifications
- Performance

Part of Speech Tagging

ILLINOIS INSTITUTE OF TECHNOLOGY

- Nonlexicalized Tagger
- Lexicalized Tagger
- Tagging Unknown Words
- K-Best Tags

Nonlexicalized Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

» Not making reference to words

» Transformation Templates :

*Change tag **a** to **b** when*

1. The preceding (following) word is tagged z.
2. The word two before (after) is tagged z.
3. One of the two preceding (following) words is tagged z.
4. One of the three preceding (following) words is tagged z.
5. The preceding word is tagged z and the following word is tagged w.
6. The preceding (following) word is tagged z and the word two before (after) is tagged w.

where a, b, z and w are variables over the set of parts of speech.

Nonlexicalized Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

```
1. apply initial-state annotator to corpus
2. while transformations can still be found do
3.     for from_tag = tag_1 to tag_n
4.       for to_tag = tag_1 to tag_n
5.         for corpus_position = 1 to corpus_size
6.           if (correct_tag(corpus_position) == to_tag
                 && current_tag(corpus_position) == from_tag)
7.                       num_good_transformations(tag(corpus_position -1))++
8.           else if (correct_tag(corpus_position) == from_tag
                 && current_tag(corpus_position) == from_tag)
9.                       num_bad_transformations(tag(corpus_position-1 ))++
10.        find max_T (num_good_transformations(T) - num_bad_transformations(T))
11.        if this is the best-scoring rule found yet then store as best rule:
             Change tag from from_tag to to_tag if previous tag is T
12.  apply best rule to training corpus
13.  append best rule to ordered list of transformations
```

Psudocode for learning transformation

Transformation Template :

Change from X to Y if the previous tag is Z

Nonlexicalized Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

	Change Tag		
#	From	To	Condition
1	NN	VB	Previous tag is *TO*
2	VBP	VB	One of the previous three tags is *MD*
3	NN	VB	One of the previous two tags is *MD*
4	VB	NN	One of the previous two tags is *DT*
5	VBD	VBN	One of the previous three tags is *VBZ*
6	VBN	VBD	Previous tag is *PRP*
7	VBN	VBD	Previous tag is *NNP*
8	VBD	VBN	Previous tag is *VBD*
9	VBP	VB	Previous tag is *TO*
10	POS	VBZ	Previous tag is *PRP*
11	VB	VBP	Previous tag is *NNS*
12	VBD	VBN	One of previous three tags is *VBP*
13	IN	WDT	One of next two tags is *VB*
14	VBD	VBN	One of previous two tags is *VB*
15	VB	VBP	Previous tag is *PRP*
16	IN	WDT	Next tag is *VBZ*
17	IN	DT	Next tag is *NN*
18	JJ	NNP	Next tag is *NNP*
19	IN	WDT	Next tag is *VBD*
20	JJR	RBR	Next tag is *JJ*

The first 20 nonlexicalized transformations
(trained on the Penn TreeBank Wall Street Journal Corpus)

Lexicalized Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

» Make reference to words

» Transformation Templates :

*Change tag **a** to **b** when*

1. The preceding (following) word is w.
2. The word two before (after) is w.
3. One of the two preceding (following) words is w.
4. The current word is w and the preceding (following) word is x.
5. The current word is w and the preceding (following) word is tagged z.
6. The current word is w.
7. The preceding (following) word is w and the preceding (following) tag is t.
8. The current word is w, the preceding (following) word is w_2 and the preceding (following) tag is t.

where w and x are variables over all words in the training corpus, and z and t are variables over all parts of speech.

Lexicalized Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

» **Example transformations** :
Change the tag …

– From **IN** to **RB** if the word two position to the right is **as**
  • … as/**RB** tall as …

– From **VBP** to **VB** if one of the previous two words is **n't**
  • We do n't eat/**VB**
  • We did n't usually drink/**VB**

Comparison of Tagging Accuracy

ILLINOIS INSTITUTE OF TECHNOLOGY

Method	Training Corpus Size (Words)	# of Rules or Context. Probs.	Acc. (%)
Stochastic	64 K	6,170	96.3
Stochastic	1 Million	10,000	96.7
Rule-Based With Lex. Rules	64 K	215	96.7
Rule-Based With Lex. Rules	600 K	447	97.2
Rule-Based w/o Lex. Rules	600 K	378	97.0

Comparison of Tagging Accuracy
(Without unknown words)

Tagging unknown words

ILLINOIS INSTITUTE OF TECHNOLOGY

» Initial-state annotator assumes the most likely tag for an unknown word is
  – "proper noun" if the word is capitalized
  – 'common noun" otherwise
» Allowable transformations

Change the tag of an unknown word(X) to Y if:

1. Deleting the prefix (suffix) x, $|x| \leq 4$, results in a word (x is any string of length 1 to 4).
2. The first (last) (1,2,3,4) characters of the word are x.
3. Adding the character string x as a prefix (suffix) results in a word ($|x| \leq 4$).
4. Word w ever appears immediately to the left (right) of the word.
5. Character z appears in the word.

Tagging unknown words

ILLINOIS INSTITUTE OF TECHNOLOGY

	Change Tag		
#	From	To	Condition
1	NN	NNS	Has suffix **-s**
2	NN	CD	Has character **.**
3	NN	JJ	Has character **-**
4	NN	VBN	Has suffix **-ed**
5	NN	VBG	Has suffix **-ing**
6	??	RB	Has suffix **-ly**
7	??	JJ	Adding suffix **-ly** results in a word.
8	NN	CD	The word **$** can appear to the left.
9	NN	JJ	Has suffix **-al**
10	NN	VB	The word **would** can appear to the left.
11	NN	CD	Has character **0**
12	NN	JJ	The word **be** can appear to the left.
13	NNS	JJ	Has suffix **-us**
14	NNS	VBZ	The word **it** can appear to the left.
15	NN	JJ	Has suffix **-ble**
16	NN	JJ	Has suffix **-ic**
17	NN	CD	Has character **1**
18	NNS	NN	Has suffix **-ss**
19	??	JJ	Deleting the prefix **un-** results in a word
20	NN	JJ	Has suffix **-ive**

The first 20 transformations for unknown words

K-Best Tagger

ILLINOIS INSTITUTE OF TECHNOLOGY

- More than one tag per word
- Rewrite rule :

  "*Add tag X to tag Y*" or "*Add tag X to word W*"

# of Rules	Accuracy	Avg. # of tags per word
0	96.5	1.00
50	96.9	1.02
100	97.4	1.04
150	97.9	1.10
200	98.4	1.19
250	99.1	1.50

Results from k-best tagger

Q & A