

Rule-Based Tagger

- The Linguistic Complaint
 - Where is the linguistic knowledge of a tagger?
 - Just a massive table of numbers
 - Aren't there any linguistic insights that could emerge from the data?
 - Could thus use handcrafted sets of rules to tag input sentences, for example, if input follows a determiner tag it as a noun.

The Brill tagger

- An example of Transformation-Based Learning
 - Basic idea: do a quick job first (using frequency), then revise it using contextual rules.
 - Painting metaphor from the readings
- Very popular (freely available, works fairly well)
- A supervised method: requires a tagged corpus

Brill Tagging: In more detail

- Start with simple (less accurate) rules...learn better ones from tagged corpus
 - Tag each word initially with most likely POS
 - Examine set of **transformations** to see which improves tagging decisions compared to tagged corpus
 - Re-tag corpus using best transformation
 - Repeat until, e.g., performance doesn't improve
 - Result: tagging procedure (ordered list of transformations) which can be applied to new, untagged text

An example

- Examples:
 - They are expected to **race** tomorrow.
 - The **race** for outer space.
- Tagging algorithm:
 1. Tag all uses of “race” as NN (most likely tag in the Brown corpus)
 - They are expected to **race/NN** tomorrow
 - the **race/NN** for outer space
 2. Use a transformation rule to replace the tag NN with VB for all uses of “race” preceded by the tag TO:
 - They are expected to **race/VB** tomorrow
 - the **race/NN** for outer space

Example Rule Transformations

		B				
S	F	r	O		Score = Fixed - Broken	
c	i	o	t		R	Fixed = num tags changed incorrect -> correct
o	x	k	h		u	Broken = num tags changed correct -> incorrect
r	e	e	e		l	Other = num tags changed incorrect -> incorrect
e	d	n	r		e	

10	18	8	14		NN -> NNP if the tag of words i+1...i+2 is 'NNP'	
9	10	1	2		NN -> VB if the tag of the preceding word is 'TO'	
8	9	1	18		NN -> VBD if the tag of the following word is 'DT'	
7	7	0	9		NN -> VBD if the tag of the preceding word is 'NNS'	
6	13	7	8		NN -> JJ if the tag of the preceding word is 'DT', and the tag of the following word is 'NN'	
7	9	2	2		NN -> NNP if the tag of the preceding word is 'NN', and the tag of the following word is ','	
8	16	8	12		NN -> NNP if the tag of words i+1...i+2 is 'NNP'	
4	6	2	11		NN -> IN if the tag of the preceding word is '.'	
3	4	1	2		NNP -> NN if the tag of words i-3...i-1 is 'JJ'	
3	3	0	2		NN -> JJ if the tag of the following word is 'JJ'	
3	3	0	4		NN -> VBP if the tag of the preceding word is 'PRP'	
3	3	0	0		WDT -> IN if the tag of the following word is 'DT'	

Sample Final Rules

Rules:

NN -> NNP if the tag of words $i+1\dots i+2$ is 'NNP'

NN -> VB if the tag of the preceding word is 'TO'

NN -> VBD if the tag of the following word is 'DT'

NN -> VBD if the tag of the preceding word is 'NNS'

NN -> JJ if the tag of the preceding word is 'DT', and the tag of the following word is 'NN'

NN -> NNP if the tag of the preceding word is 'NN', and the tag of the following word is ','

NN -> NNP if the tag of words $i+1\dots i+2$ is 'NNP'

NN -> IN if the tag of the preceding word is '.'

NNP -> NN if the tag of words $i-3\dots i-1$ is 'JJ'

NN -> JJ if the tag of the following word is 'JJ'

NN -> VBP if the tag of the preceding word is 'PRP'

WDT -> IN if the tag of the following word is 'DT'

NN -> JJ if the tag of the preceding word is 'IN', and the tag of the following word is 'NN'

NN -> VBN if the tag of the preceding word is 'VBP'

VBD -> VB if the tag of the preceding word is 'MD'

NN -> JJ if the tag of the preceding word is 'CC', and the tag of the following word is 'NN'