Neural networks

Natural language processing - word tagging

WORD TAGGING

Topics: word tagging

- In many NLP applications, it is useful to augment text data with syntactic and semantic information
  - we would like to add syntactic/semantic labels to each word
- This problem can be tackled using a conditional random field with neural network unary potentials
  - we will describe the model developed by Ronan Collobert and Jason Weston in:

    A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning
    Collobert and Weston, 2008

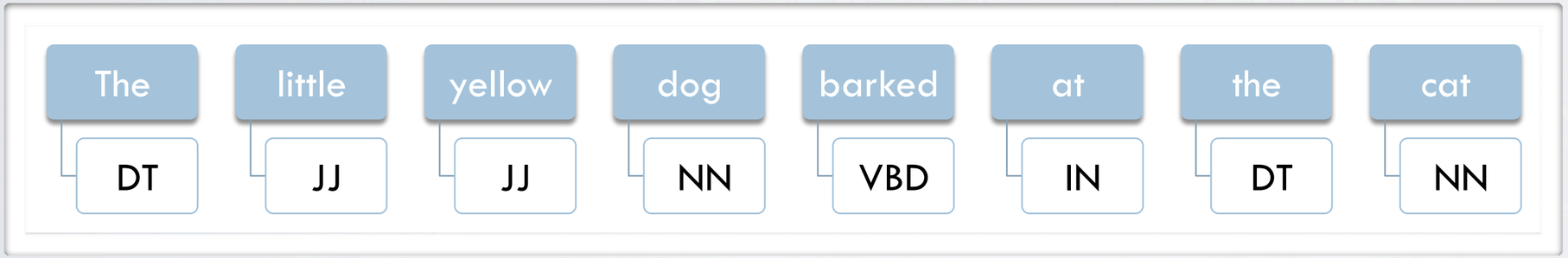
    (see Natural Language Processing (Almost) from Scratch for the journal version)

WORD TAGGING

Topics: part-of-speech tagging

- Tag each word with its part of speech category
  - noun, verb, adverb, etc.
  - might want to distinguish between singular/plural, present tense/past tense, etc.
  - see Penn Treebank POS tags set for an example

- Example:

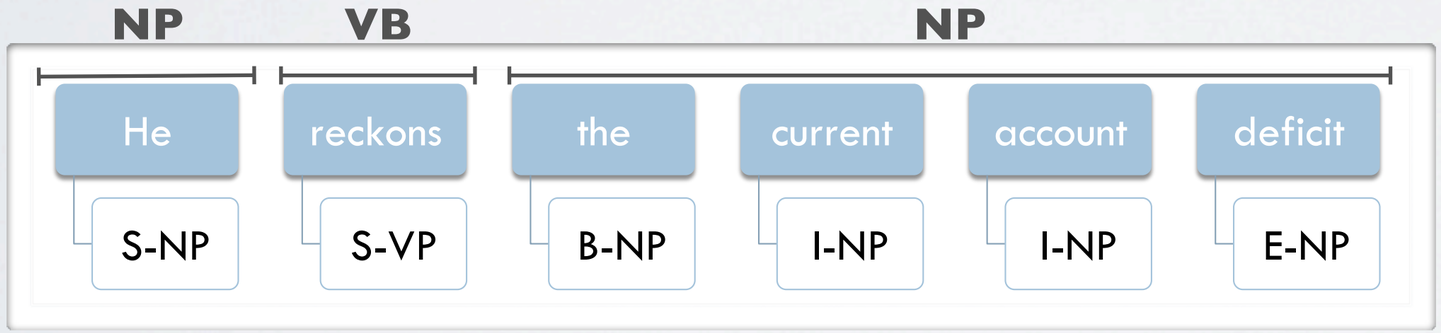


(from Stanislas Lauly)

WORD TAGGING

Topics: chunking

- Segment phrases into syntactic phrases
  - noun phrase, verb phrase, etc.
- Segments are identified with IOBES encoding
  - single word phrase (`S-` prefix). Ex.: `S-NP`
  - multiword phrase (`B-`, `I-`, `E-` prefixes). Ex.: `B-VP I-VP I-VP E-VP`
  - words outside of syntactic phrases: `O`



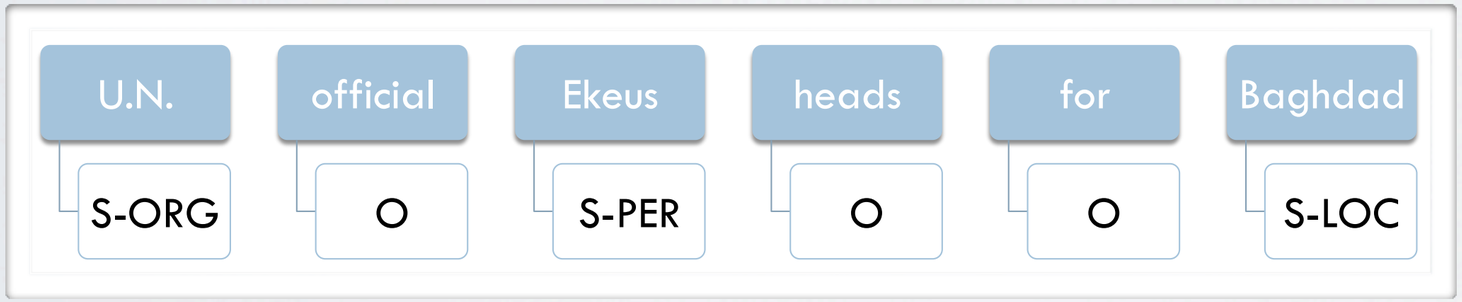
(from Stanislas Lauly)

WORD TAGGING

Topics: named entity recognition (NER)

- Identify phrases referring to a named entity
  - person
  - location
  - organization

- Example:

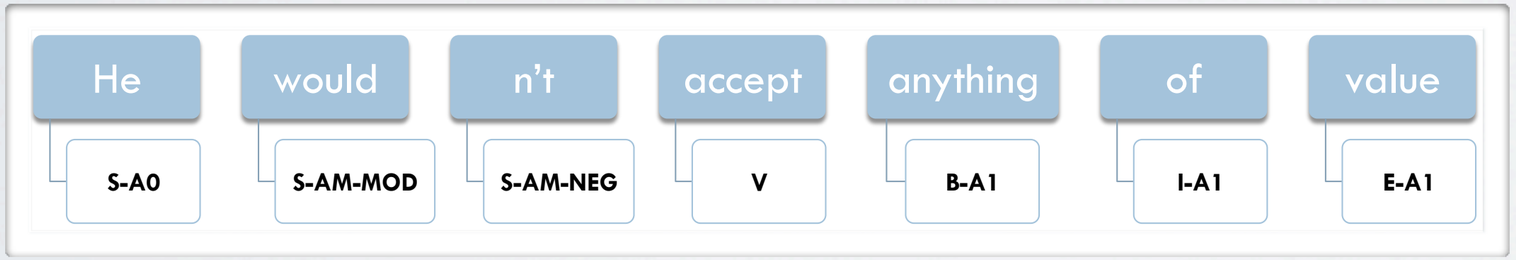


(from Stanislas Lauly)

WORD TAGGING

Topics: semantic role labeling (SRL)

- For each verb, identify the role of other words with respect to that verb
- Example:
  - `V`: verb
  - `A0`: acceptor
  - `A1`: thing accepted
  - `A2`: accepted from
  - `A3`: attribute
  - `AM-MOD`: modal
  - `AM-NEG`: negation



(from Stanislas Lauly)

WORD TAGGING

Topics: labeled corpus

- The raw data looks like this:

The	DT	B-NP	O	B-A0	B-A0
$	$	I-NP	O	I-A0	I-A0
1.4	CD	I-NP	O	I-A0	I-A0
billion	CD	I-NP	O	I-A0	I-A0
robot	NN	I-NP	O	I-A0	I-A0
spacecraft	NN	E-NP	O	E-A0	E-A0
faces	VBZ	S-VP	O	S-V	O
a	DT	B-NP	O	B-A1	O
six-year	JJ	I-NP	O	I-A1	O
journey	NN	E-NP	O	I-A1	O
to	TO	B-VP	O	I-A1	O
explore	VB	E-VP	O	I-A1	S-V
Jupiter	NNP	S-NP	S-ORG	I-A1	B-A1
and	CC	O	O	I-A1	I-A1
its	PRP$	B-NP	O	I-A1	I-A1
16	CD	I-NP	O	I-A1	I-A1
known	JJ	I-NP	O	I-A1	I-A1
moons	NNS	E-NP	O	E-A1	E-A1
.	.	O	O	O	O