

LING439/539 - Statistical NLP

Chapter 3. Linguistic essentials

Tuesday, August 30, 2016

This chapter introduces basic linguistic concepts...

Part of speech and morphology

Word classes show similar syntactic behavior, and often a typical semantic type.

- ▶ Syntactic or grammatical categories
- ▶ Parts of speech (POS).

Words classes are normally divided into two:

- ▶ Open or lexical categories (noun, verbs, and adjectives)
 - ▶ new words are commonly added
- ▶ Closed or functional categories (prepositions and determiners)
 - ▶ there are only a few members, and the members of which normally have a clear grammatical use.

Pierre Vinken , 61 years old , will

join the board as a nonexecutive director .

See about Mr. Pierre Vinken:

<http://languagelog.ldc.upenn.edu/n11/?p=3594>

Nouns (and pronouns)

Nouns typically refer to entities in the world like people, animals, and things.

Pierre Vinken , 61 years old , will
NNP NNP NNS

join the board as a nonexecutive director .
NN NN

NN Noun, singular or mass

NNS Noun, plural

NNP Proper noun, singular

NNPS Proper noun, plural

Determiners and adjectives

Determiners describe the particular reference of a noun and adjectives are used to describe properties of nouns.

Pierre Vinken , 61 years old , will
JJ

join the board as a nonexecutive director .
DT DT JJ

JJ Adjective

JJR Adjective, comparative

JJS Adjective, superlative

...

DT Determiner

Verbs

Verbs are used to describe actions and states.

Pierre Vinken , 61 years old , will
MD

join the board as a nonexecutive director .
VB

VB Verb, base form

VBD Verb, past tense

VBG Verb, gerund or present participle

...

MD Modal auxiliaries

Other parts of speech

Adverbs specify place, time, manner or degree:

- ▶ She *often* travels to Las Vegas.
- ▶ She *allegedly* committed perjury.
- ▶ She started her career off very *impressively*.

Prepositions and particles:

- ▶ She ran up/**IN** a hill.
- ▶ She ran up/**RP** a bill.

Note the meaning shift from the literal meaning of running on an incline in *ran up/**IN** a hill* to the figurative meaning of building up a large bill in *ran up/**RP** a bill..*

Phrase structure

Certain grouping of words behave as *constituents*.

- ▶ phrase structure grammars, or
- ▶ constituent grammars

Noun phrases

- ▶ The noun is the **head** of the noun phrase
- ▶ usually the arguments of the verbs
- ▶ participants in the action, activity of state described by the verb.

[_{NP} Pierre Vinken] will join [_{NP} the board] as [_{NP} a nonexecutive director]

Prepositional phrases

- ▶ The noun is headed by a preposition and contain a noun phrase complement

Pierre Vinken will join the board $\left[_{PP} \text{ as } \left[_{NP} \text{ a nonexecutive director} \right] \right]$

Verb phrases

Pierre Vinken

$$\left[\begin{array}{l} \text{VP will} \\ \left[\begin{array}{l} \text{VP join } [\text{NP the board}] \left[\text{PP as } [\text{NP a nonexecutive director}] \right] \end{array} \right] \end{array} \right]$$

Putting together

In the Penn Treebank:

Phrase structure grammars

- ▶ $S \rightarrow NP VP$
- ▶ $NP \rightarrow NNP NNP$
- ▶ $VP \rightarrow MD VP$
- ▶ ...
- ▶ $NN \rightarrow \textit{director}$

Subcategorization

She	<i>put</i>	the book	on the table
NP		NP	PP

→ *put* requires NP, NP, and PP.

NP, NP, and PP for *put* are **arguments**, which express entities that are centrally involved in the activities of the verb (subject, object, direct object, etc.)

He *gave* his presentation on the stage
NP NP PP

→ *gave* requires NP, NP, and PP ?

[_{PP} *on the stage*] is optional and an **adjunct**.

Phrase structure ambiguity

Daisuke Kawahara and Sadao Kurohashi (2007). Probabilistic Coordination Disambiguation in a Fully-Lexicalized Japanese Parser. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.