Word Classes

Parts of Speech in Natural Language Processing

Study Notes: Definition, Important Points, Architecture, Example and Applications

1. Definition

Word classes (also called **parts of speech**, PoS, or **lexical categories**) are groups of words that behave alike grammatically. Words in the same class occur in similar positions in sentences, take similar affixes, and usually play a similar role in meaning. Traditional grammar names eight classes: noun, pronoun, verb, adjective, adverb, preposition, conjunction and interjection. Modern NLP uses finer tagsets.

> **Example:** *dog, city, happiness* are all **nouns**: each can follow “the” (the dog), take a plural -s (dogs) and act as the subject of a verb.

2. Important Points

2.1 Criteria used to assign a word class

Criterion	Idea	Example
Morphological	Which affixes the word can take.	Nouns take -s (plural); verbs take -ed (past), -ing; adverbs often end in -ly.
Syntactic / distributional	Which words can appear before or after it.	Any word fitting “the ___ is here” is a noun.
Semantic (notional)	General meaning type. Least reliable, since there are many exceptions.	Nouns name things, verbs name actions, adjectives name properties. But “destruction” is a noun naming an action.

2.2 Open class vs closed class

Feature	Open class	Closed class
Membership	Large and growing; new words are added easily	Small and fixed; new words are rarely added
Also called	Content words (lexical words)	Function words (grammatical words)
Members	Nouns, verbs, adjectives, adverbs	Determiners, pronouns, prepositions, conjunctions, particles, numerals, auxiliaries
Frequency	Many different words, each fairly rare	Few words, each very frequent (the, of, and, a, to)
New-word example	selfie, google (verb), blockchain	No new pronouns or prepositions appear easily

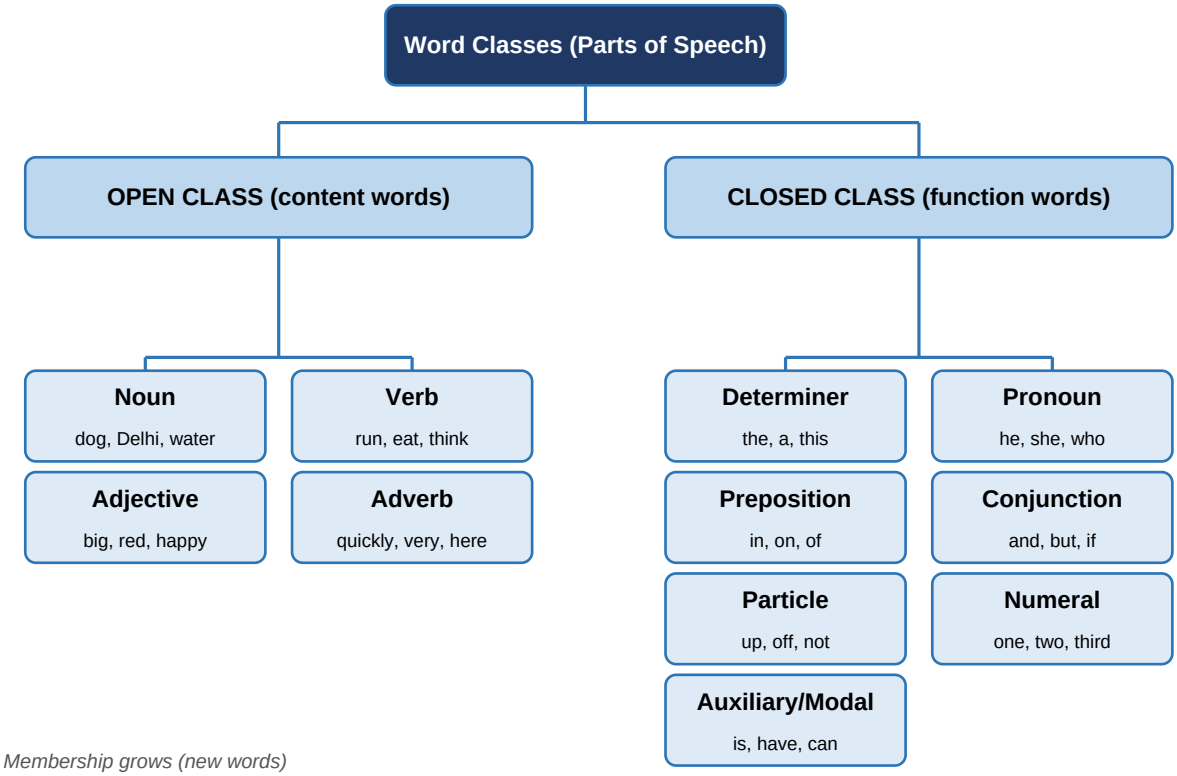
2.3 Other key points

- **Four major open classes:** noun, verb, adjective, adverb. Nouns are subdivided into proper (*Hyderabad*) and common (count: *book*; mass: *water*). Verbs are subdivided into main verbs, auxiliaries (*have, be*) and modals (*can, will*).

- **Ambiguity:** one word can belong to several classes. Only a minority of word types are ambiguous, but they include very common words, so a large share of running text (about 40% or more of tokens) is ambiguous. Example: *back* can be an adjective, noun, adverb, verb or particle.
- **Tagsets:** the Brown corpus tagset has 87 tags, the Penn Treebank tagset has 45, and Universal Dependencies has 17 universal tags (NOUN, VERB, ADJ, ADV, PRON, DET, ADP, CCONJ, ...).
- **Language dependence:** English uses prepositions (*in the house*), whereas Hindi and Telugu use postpositions placed after the noun, so word-class inventories differ across languages.
- **Class determines behaviour:** the word class decides a word's inflection, its place in syntax and even its pronunciation (*REcord* noun vs *reCORD* verb).
- **Tag set granularity:** a finer tagset gives more information (VBD vs VBZ) but is harder to predict automatically.

3. Architecture

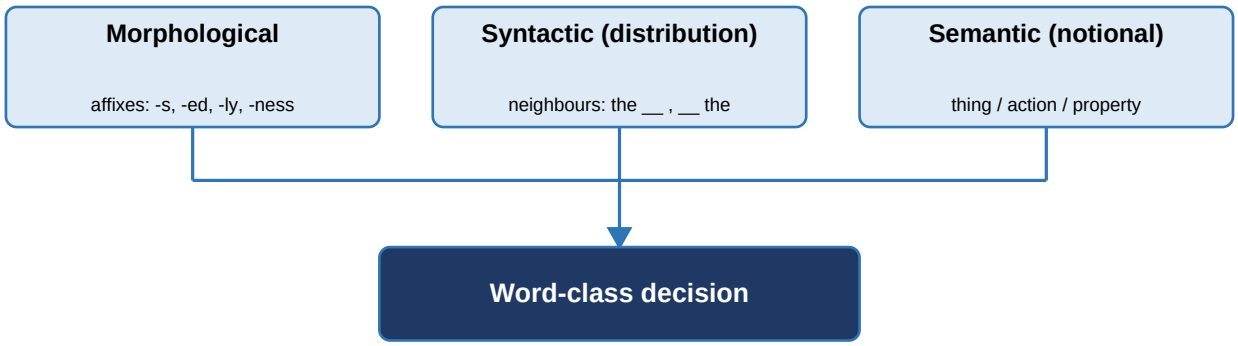
3.1 Classification hierarchy of word classes



Membership grows (new words)

3.2 How a word is assigned to a class

A word is placed in a class by combining morphological, syntactic and semantic evidence. In a computer system the same evidence is used by the PoS tagger.



3.3 Common English word classes with Penn Treebank tags

Class	Type	Penn tags	Examples
Noun	Open	NN, NNS, NNP, NNPS	cat, cats, Ravi, Himalayas
Verb	Open	VB, VBD, VBG, VBN, VBP, VBZ	eat, ate, eating, eaten, eats
Adjective	Open	JJ, JJR, JJS	tall, taller, tallest
Adverb	Open	RB, RBR, RBS	quickly, very, soon
Determiner	Closed	DT	the, a, an, this, those
Pronoun	Closed	PRP, PRP$, WP	he, she, my, who
Preposition	Closed	IN	in, on, of, with, from
Conjunction	Closed	CC, IN	and, but, or, because
Modal / Auxiliary	Closed	MD, VB*	can, will, should, have, be
Particle	Closed	RP	up, off, out (give up)
Numeral	Closed	CD	one, 25, three, 2026

4. Example

4.1 Tagging every word of a sentence with its class

Sentence: **“The quick brown fox quickly jumped over three lazy dogs and she laughed.”**

Word	Tag	Word class	Open / Closed	Reason
The	DT	Determiner	Closed	Precedes a noun phrase
quick	JJ	Adjective	Open	Modifies the noun “fox”
brown	JJ	Adjective	Open	Modifies the noun “fox”
fox	NN	Noun	Open	Head of the noun phrase; can take -es
quickly	RB	Adverb	Open	Ends in -ly, modifies the verb
jumped	VBD	Verb (past)	Open	Takes -ed; denotes an action
over	IN	Preposition	Closed	Introduces the phrase “three lazy dogs”
three	CD	Numeral	Closed	Quantifies “dogs”
lazy	JJ	Adjective	Open	Describes “dogs”
dogs	NNS	Noun (plural)	Open	Plural -s
and	CC	Conjunction	Closed	Joins two clauses
she	PRP	Pronoun	Closed	Replaces a noun as subject
laughed	VBD	Verb (past)	Open	Takes -ed; denotes an action

Result: The/DT quick/JJ brown/JJ fox/NN quickly/RB jumped/VBD over/IN three/CD lazy/JJ dogs/NNS and/CC she/PRP laughed/VBD

4.2 Substitution test (distributional test)

Frame: “I saw the ___ yesterday.” Words that fit the same frame belong to the same class.

Word	Fits the frame?	Conclusion
dog, book, teacher	Yes	Noun
run, quickly	No	Not a noun (verb, adverb)
beautiful	No (needs a noun after it)	Adjective

4.3 Same word, different class (ambiguity)

Sentence	Word	Class
I read a **book**.	book	Noun (NN)
Please **book** a ticket.	book	Verb (VB)
She went **back**.	back	Adverb (RB)
He sat at the **back**.	back	Noun (NN)
The **back** door was open.	back	Adjective (JJ)

5. Applications

- **PoS tagging:** word classes are the label set that taggers (rule-based, HMM, MaxEnt, neural) predict.
- **Syntactic parsing:** grammar rules such as NP → DT JJ NN are written in terms of word classes.
- **Information retrieval and search:** closed-class words are removed as stop words; nouns and noun phrases are kept as index terms.
- **Named entity recognition:** proper nouns (NNP) are strong clues for names of people, places and organizations.
- **Text-to-speech:** the class decides pronunciation (*record* noun vs verb; *present* noun vs verb).
- **Machine translation:** word-class patterns drive reordering, for example English adjective + noun versus noun + adjective in French.
- **Lemmatization and stemming:** the correct base form depends on the class (*saw* as noun → saw; as verb → see).
- **Sentiment analysis:** adjectives and adverbs (*excellent, poor, very*) carry most opinion information.
- **Word-sense disambiguation and question answering:** knowing the class narrows the possible meanings of a word.
- **Grammar and spell checkers, language teaching tools:** detect wrong word-class use such as “He runs quick” (should be “quickly”).

Summary: Word classes group words by shared grammatical behaviour. They split into open classes (content words that keep growing) and closed classes (a fixed set of function words). They are the foundation for PoS tagging, parsing and almost every higher-level NLP task.