

Lemmatization

- Whereas stemming simply removes suffixes, lemmatization reduces inflected words to a semantically correct base word, called a *lemma*, that lies in the dictionary of valid words.

goose -> goose

geese -> goose

- Whereas stemmers are algorithmic in nature, lemmatizers require data (memory) in order to evaluate semantic relationships between words.
- Common lemmatized used the English language is the WordNet Lemmatizer.

Part of Speech (POS) tagging

- POS tagging is the task of assigning tags to each word in a sentence according to its part of speech.

The	quick	brown	fox	jumped	over	the	lazy	dog
DET	VERB	VERB	NOUN	VERB	PREP	DET	VERB	NOUN

- The set of tags, or *Parts*, varies by dataset, task etc., and can be much larger than the standard set of eight.
- Used in syntactic parsing, and for word sense disambiguation (WSD)

leaves -> leave
VERB

leaves -> leaf
NOUN