

Meanings in Human- Annotated Dictionaries



Goals

1. Familiarising with WordNet, one of the most popular human-annotated datasets
2. Clarifying the scope of application of WordNet in NLP systems



**Why do
human-
annotated
dictionaries
matter?**



Human-annotated dictionaries are key inputs for pattern-matching

Human beings are entrenched in the symbols and social categories of natural language, while machines are not. Hence, machines are not able to associate meanings with lexemes

Human-annotated dictionaries provide machines with 'pattern-matching' rules that induce meaningful responses on the part of an NLP system (e.g., a chatbot) *vis à vis* natural language inputs (e.g., a customer raising questions in a chat)



WordNet



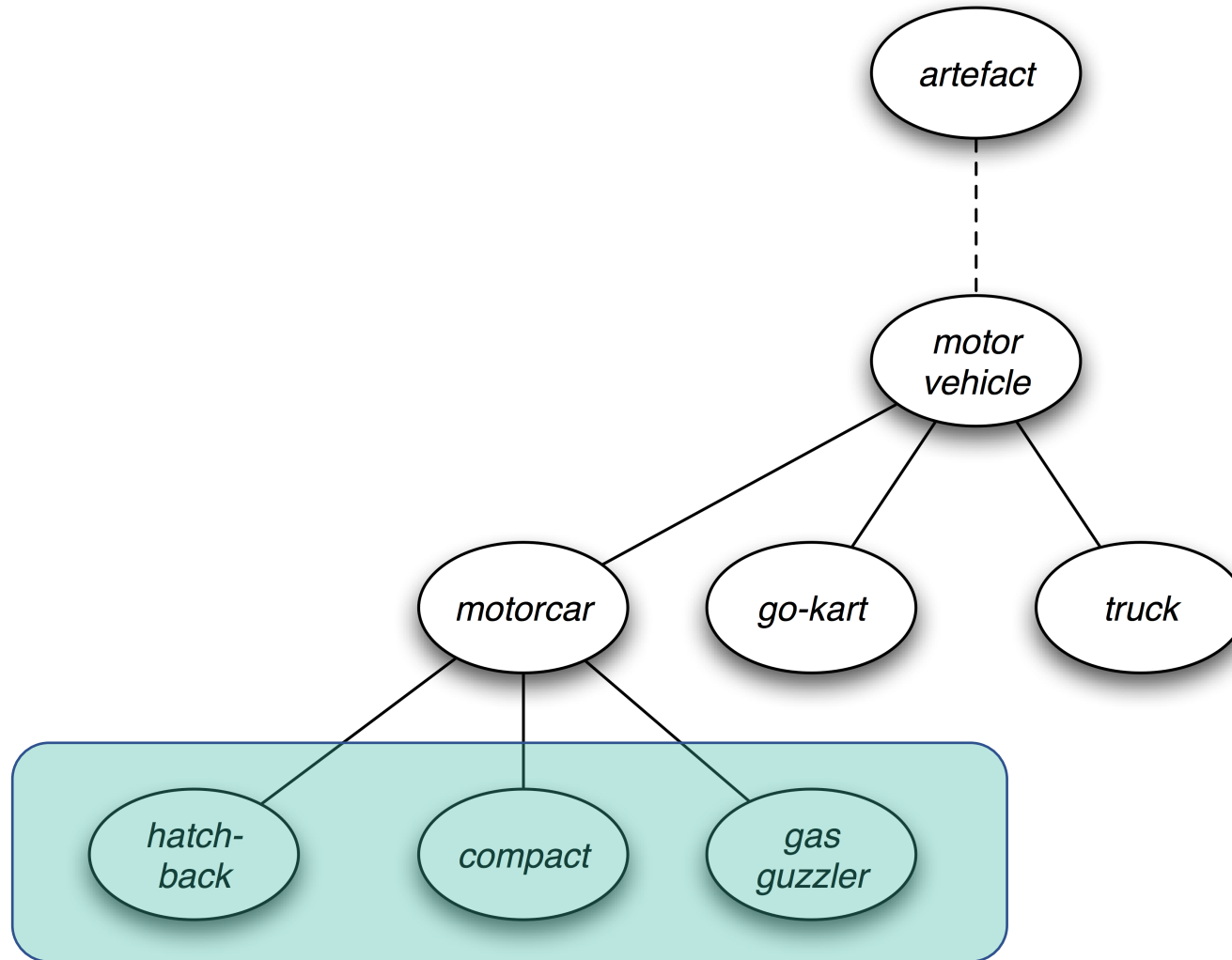
WordNet — the most popular human-annotated dictionary

[WordNet®](#) is an extensive lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into homogenous sets, expressing a distinct concept

WordNet superficially resembles a thesaurus in that it groups words based on their meanings. However, Wordnet presents some critical nuances (see graph on the right)



A fragment of the WordNet concept hierarchy



The structure of WordNet

The primary relation among words in WordNet is synonymy, as between the words 'shut' and 'close' or 'car' and 'automobile'. Synonyms — terms that denote the same concept and are interchangeable in many contexts — are grouped into unordered sets (synsets)

WordNet's 117 000 synsets are linked to other synsets using a small number of 'conceptual relations'

Additionally, a synset contains a brief definition ('gloss') and, in most cases, one or more short sentences illustrating the use of the synset members

Word forms with several distinct meanings are represented in as many distinct synsets



The relationships included in WordNet

The most frequently encoded relation among synsets is the super-subordinate relation (also called hyperonymy, hyponymy or ISA relation)

It links more general synsets like 'furniture' to increasingly specific ones like 'bed' and 'bunkbed'

Thus, WordNet states that the category furniture includes bed, which includes bunkbed; conversely, concepts like bed and bunkbed make up the category furniture

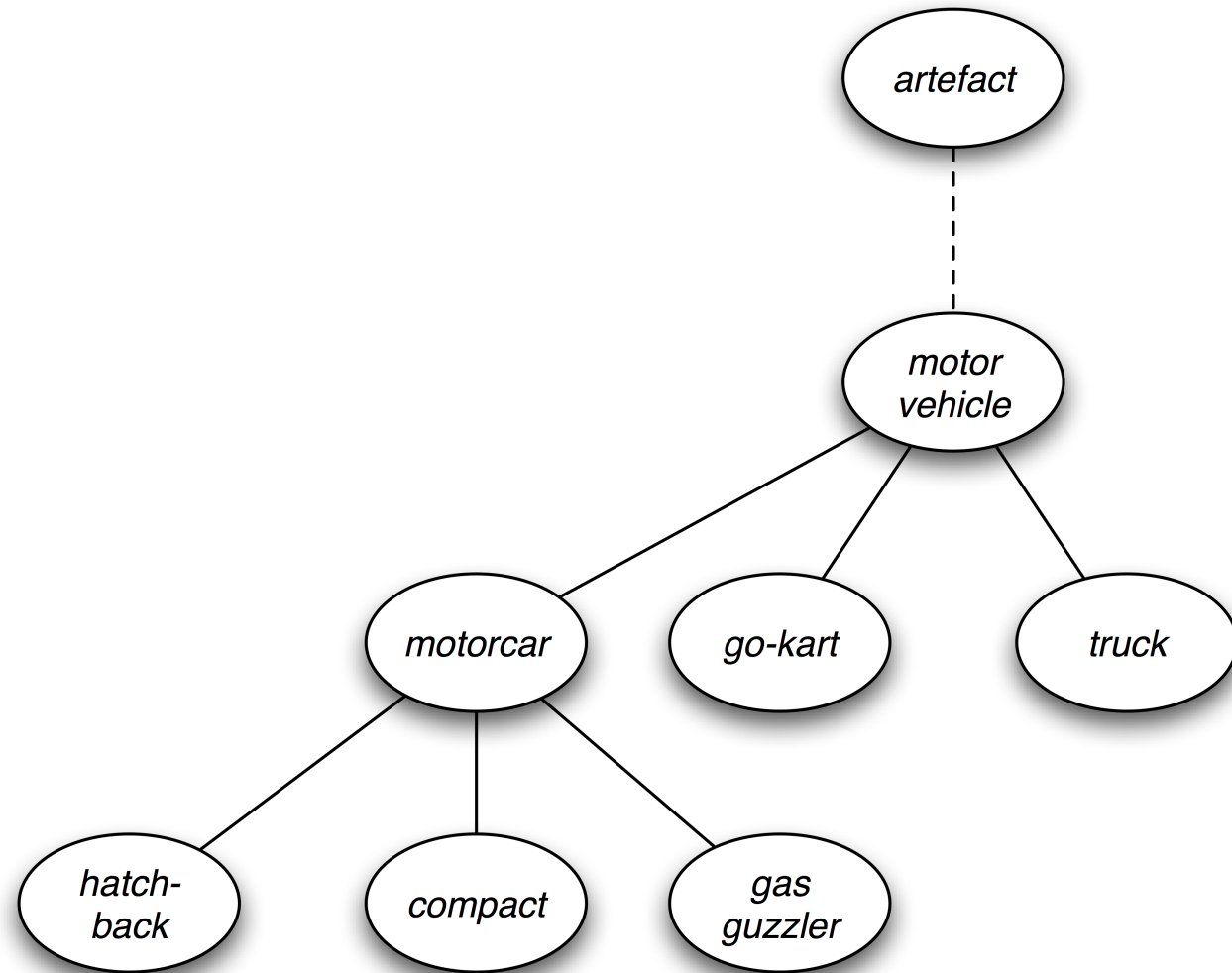


A fragment of the WordNet concept hierarchy

Abstract concepts



Concrete terms



NLP systems drawing on WordNet

WordNet can be used to create or enhance many NLP systems, for example, chatbots

Recommendation systems are one of the most common applications of WordNet. For example, synsets are particularly useful to provide customers with bundles of suggested items. Given that a customer has purchased a lamp, we may want to recommend an item in the same synset, for example, a complementary product such as light bulbs



So, is WordNet 'enough' to do NLP?



- It bring usable meanings to machines (e.g., it can inform our chat-bot)
- Great as a resource for research & teaching



- Bottleneck: WordNet is an annotated dataset
- It requires human labor to adapt
 - perhaps, it's impossible to keep-up-to date
 - at least, it misses new meanings of words
- It is subjective
- It misses missing nuance (Manning, 2019): e.g. 'proficient' is listed as a synonym for 'good'
- It doesn't offer a continuous measure of word similarity



Wrap-up



Main points

- A WordNet is a human-annotated dictionary that groups words into homogenous sets (synsets)
- At the same time, synsets are hierarchically organised via the hyponymy - hyperonymy relationship
- WordNet can support various NLP systems that involve using the meanings human beings give to words and concepts

