

CS11-747 Neural Networks for NLP

Models of Words

Graham Neubig



Carnegie Mellon University

Language Technologies Institute

Site

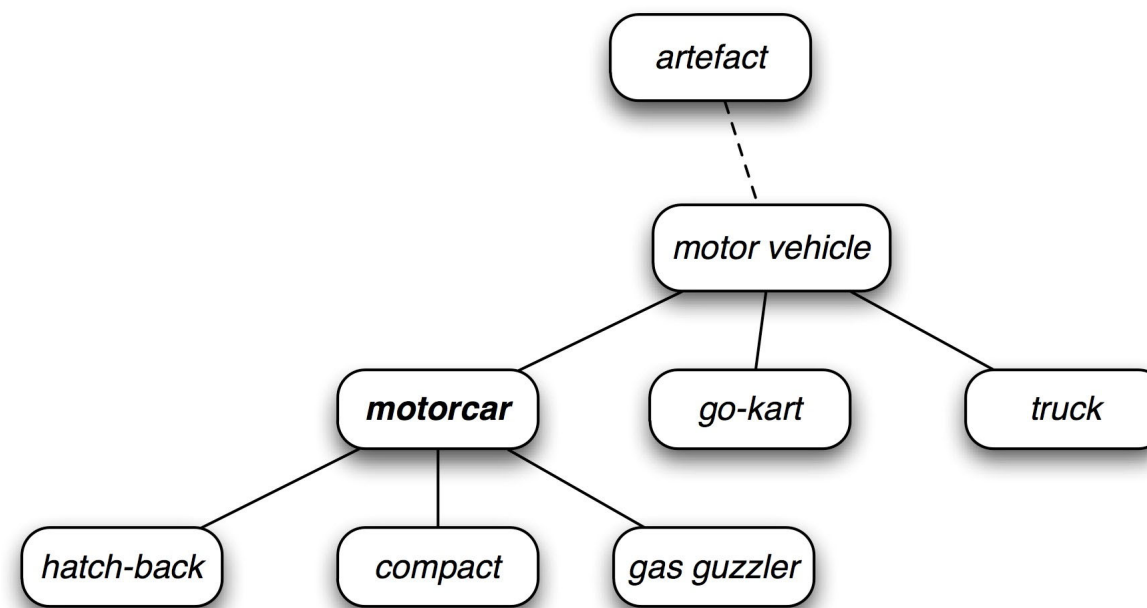
[h](https://www.cmu.edu/language-technologies/institute/human-computer-interaction/teaching/cs11-747-2019/) [/class/nn4nlp2019/](https://www.cmu.edu/language-technologies/institute/human-computer-interaction/teaching/cs11-747-2019/)

What do we want to know about words?

- Are they the same part of speech?
- Do they have the same conjugation?
- Do these two words mean the same thing?
- Do they have some semantic relation (is-a, part-of, went-to-school-at)?

A Manual Attempt: WordNet

- WordNet is a large database of words including parts of speech, semantic relations



- Major effort to develop, projects in many languages.
- But can we do something similar, more complete, and without the effort?

An Answer (?): Word Embeddings!

- A continuous vector representation of words

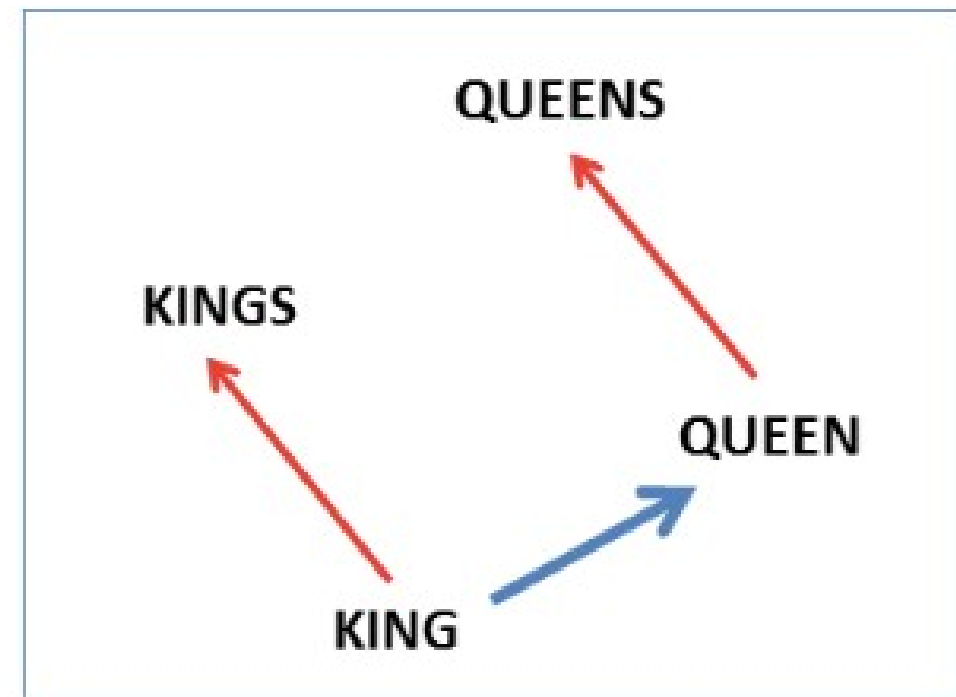
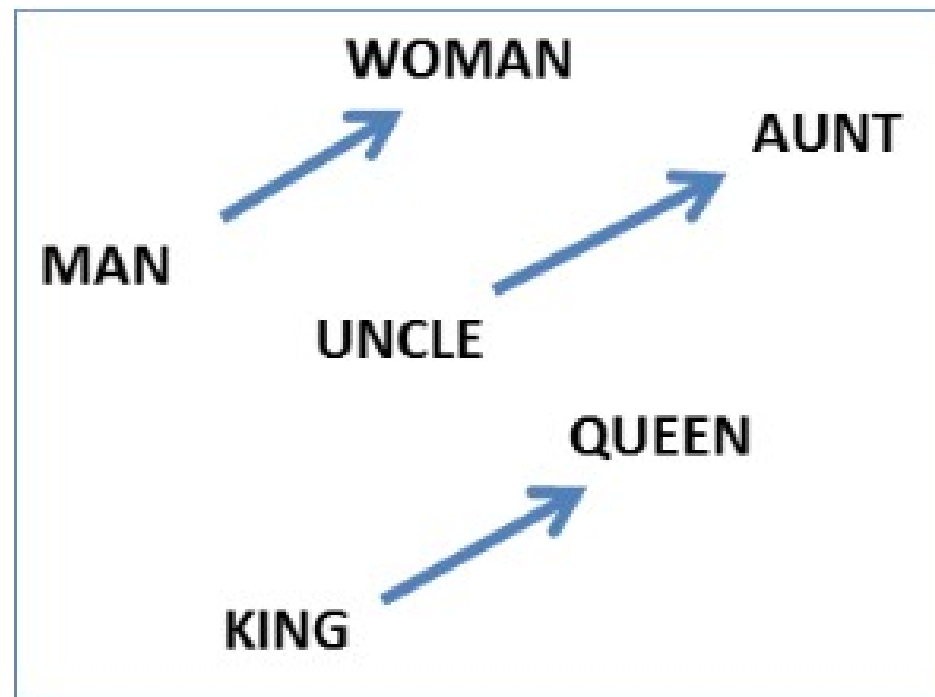


- Within the word embedding, these features of syntax and semantics may be included
 - Element 1 might be **more positive for nouns**
 - Element 2 might be **positive for animate objects**
 - Element 3 might have **no intuitive meaning whatsoever**

Word Embeddings are Cool!

(An Obligatory Slide)

- e.g. king-man+woman = queen (Mikolov et al. 2013)



- “What is the female equivalent of king?” is not easily accessible in many traditional resources

How to Train Word Embeddings?

- **Initialize randomly**, train jointly with the task
- Pre-train on a **supervised** task (e.g. POS tagging) and test on another, (e.g. parsing)
- Pre-train on an **unsupervised** task (e.g. `word2vec`)

Unsupervised Pre-training of Word Embeddings

(Summary of Goldberg 10.4)

Distributional vs. Distributed Representations

- **Distributional representations**
 - Words are similar if they appear in similar contexts (Harris 1954); distribution of words indicative of usage
 - In contrast: *non-distributional* representations created from lexical resources such as WordNet, etc.
- **Distributed representations**
 - Basically, something is represented by a vector of values, each representing activations
 - In contrast: *local* representations, where represented by a discrete symbol (one-hot vector)

Distributional Representations

(see Goldberg 10.4.1)

- **Words** appear in a **context**

<s>	<s>	<unk>	communications	pittsburgh	acquired	<unk>	&	co.
investment	management	inc.	a	pittsburgh	firm	that	runs	a
<s>	mr.	allen	's	pittsburgh	firm	advanced	investment	management
look	stupid	<unk>	former	pittsburgh	<unk>	second	<unk>	<unk>
through	the	university	of	pittsburgh	law	school	<s>	<s>
with	the	university	of	pittsburgh	<s>	<s>	<s>	<s>
<unk>	he	heads	the	pittsburgh	branch	of	the	committee
at	the	university	of	pittsburgh	earn	up	to	\$
for	society	corp.	a	cleveland	bank	said	demand	for
as	washington	<unk>	r.i.	cleveland	<unk>	n.c.	minneapolis	and
<s>	<s>	<unk>	a	cleveland	merchant	bank	owns	about
new	stadiums	ranging	from	cleveland	to	san	antonio	and
<s>	the	philadelphia	and	cleveland	districts	for	example	reported
mcdonald	&	co.	in	cleveland	said	<unk>	's	unanticipated
<unk>	tumor	at	the	cleveland	clinic	in	N	<s>
at	mcdonald	&	co.	cleveland	<s>	<s>	<s>	<s>

(try it yourself w/ `kwic.py`)

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/127122146145006025>