Auditory behaviour, Language and Sentiment Analysis

DSA8022 Frontiers in Analytics

Dr Gary McKeown
School of Psychology, David Keir Building
0G3.536

Auditory behaviour, Language and Sentiment Analysis

DSA8022 Frontiers in Analytics

Dr Gary McKeown

School of Psychology, David Keir Building

0G3.536

OVERVIEW

- Nature of Language
- Natural Language Processing
- Latent Semantic Analysis - Vector Space models
- Topic Models
- Sentiment Analysis
- Paralanguage - Opensmile Praat
- Transcription

HUMAN COMMUNICATION

Towards the symbolic

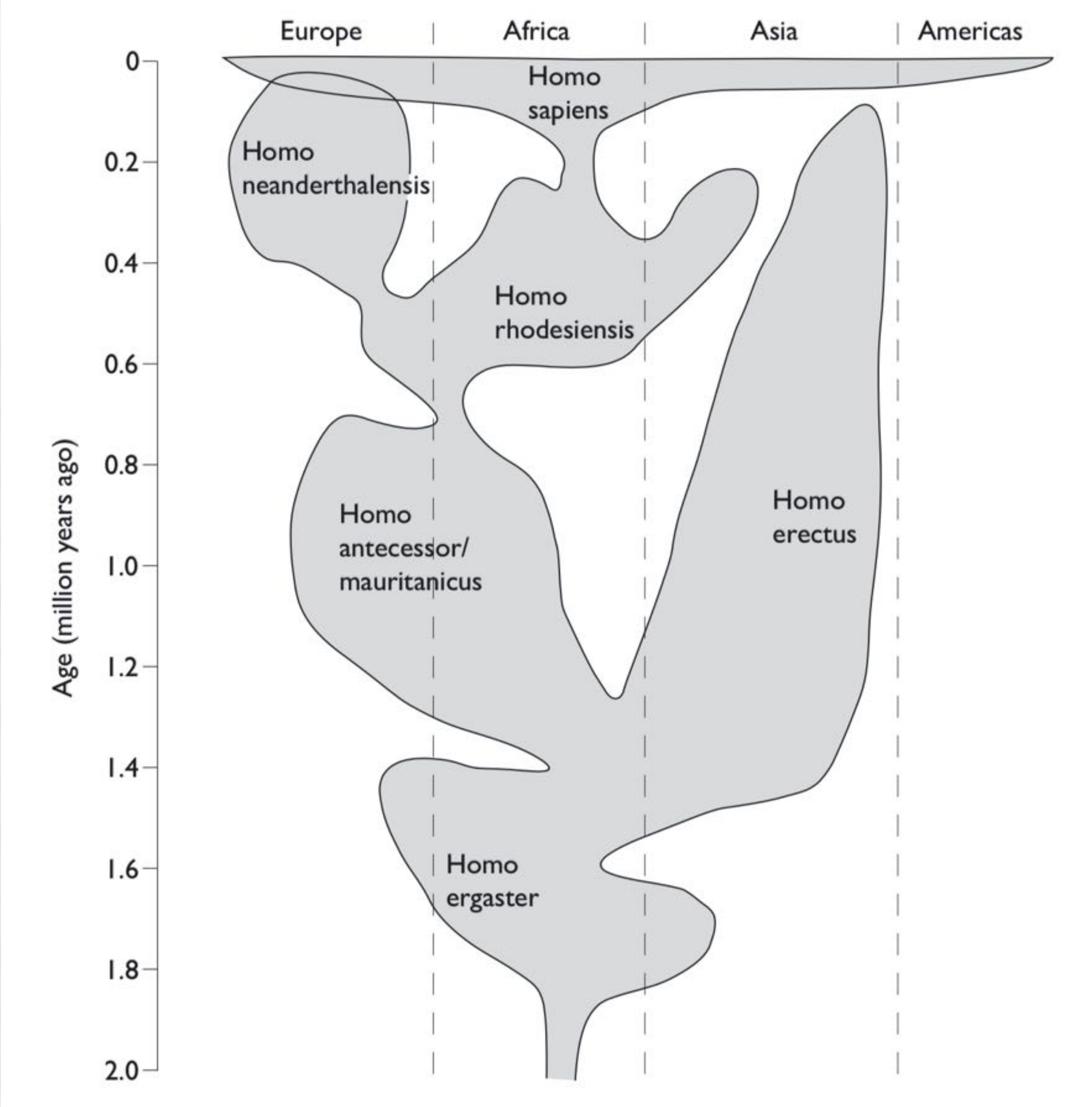
EVOLUTION OF LANGUAGE

- Now a big topic in cognitive science
  - Evolang 1, 1996 - Evolang Toruń 2018 (Poland) 12th
    - James Hurford
  - Classic papers:
    - Pinker & Bloom (1990)
    - Hauser, Chomsky & Finch (2002)
    - Kirby, Cornish, & Smith (2008)

EVOLUTION OF LANGUAGE

- Language did not evolve in a vacuum
  - Existing communication - non-verbal
  - parallel with facial expression, gesture, paralanguage, head movements.
- Relevance theory - lots of underdetermined components
- Meaning is a combined weighted probability of each of these aspects - lots of little bits of evidence
  - although symbolic language is one of the most important

EVOLUTION OF LANGUAGE



NATIVISTS AND EMPIRICISTS

WHAT IS LANGUAGE?

- "A system of rules and symbols that enable us to communicate" Harley(2001)
- Animal language
- Body language

ELEMENTS OF LANGUAGE

- Words
  - Mental Lexicon
- Morphology
  - Morphemes
    - Word parts, suffixes, prefixes
- Syntax
  - Grammar, parsing
- Semantics
  - Meanings of words
  - Semantic representations
- Pragmatics
  - Context and environment

NATIVIST VS EMPIRICIST

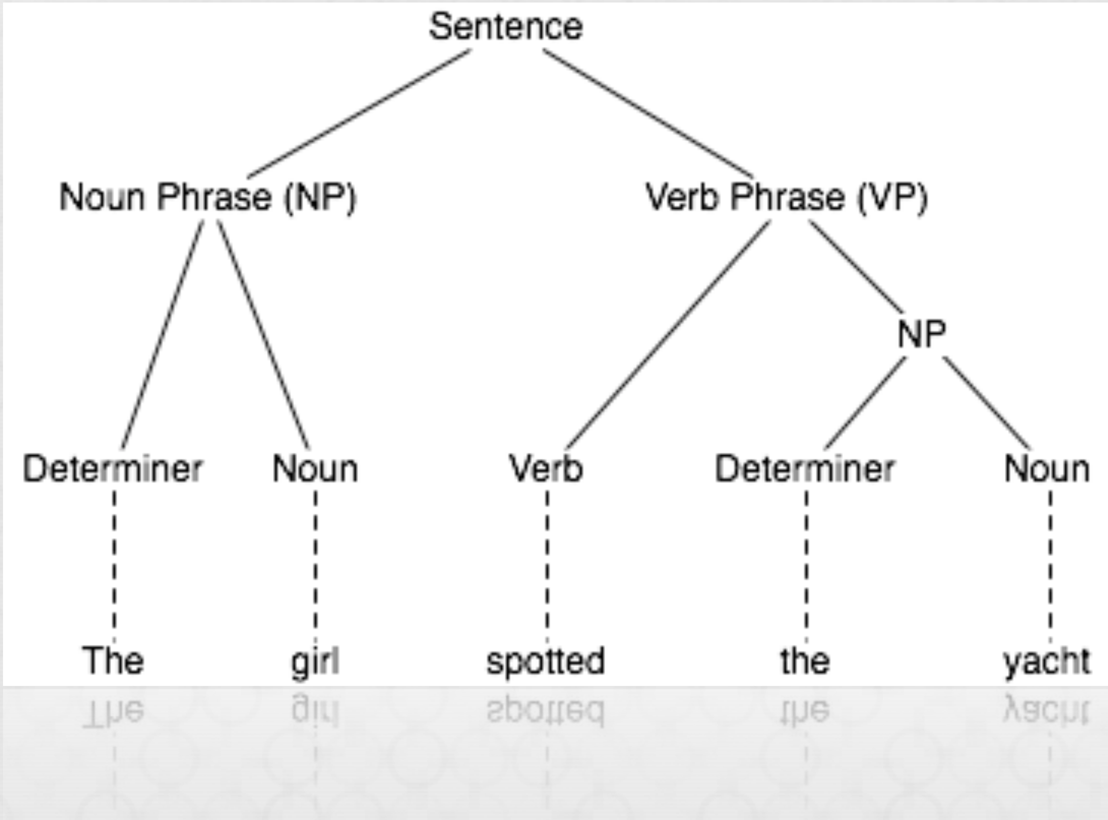
- Linguistic Nativist (Nature)
  - Chomsky, Fodor, Pinker,
    - Innate, Rules, modular, domain-specific.
- Linguistic Empiricist (Nurture)
  - Rumelhart & McClelland, Connectionism
    - Environment, statistical, graded, domain-general

LINGUISTIC NATIVIST VIEW

- Chomsky, Pinker - the language instinct
- Innate - genetically determined
- Rule guided processing
- Computational model of mind
- Rule guided transformation of structured mental representations

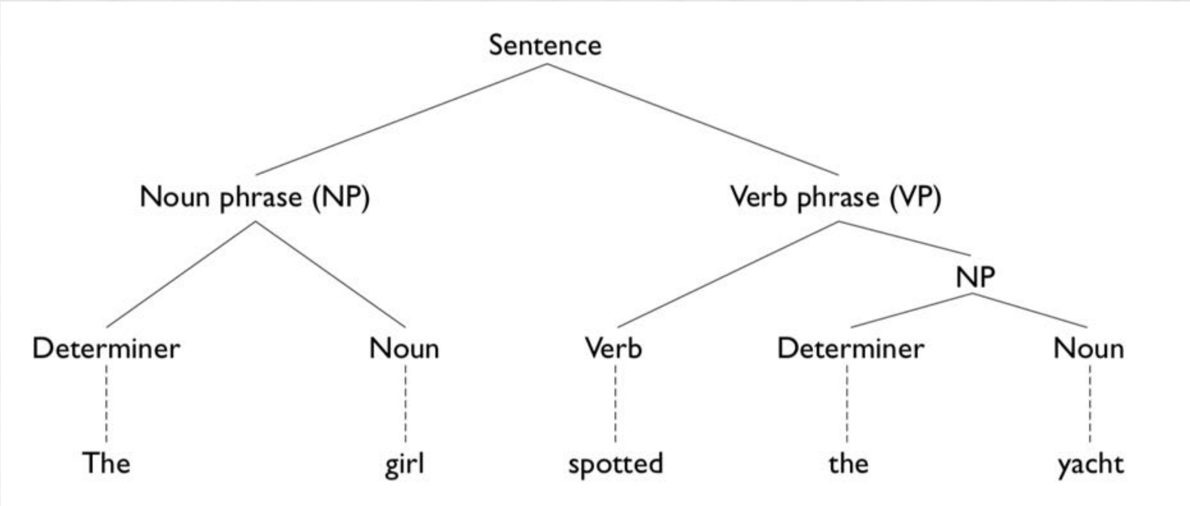
NOAM CHOMSKY

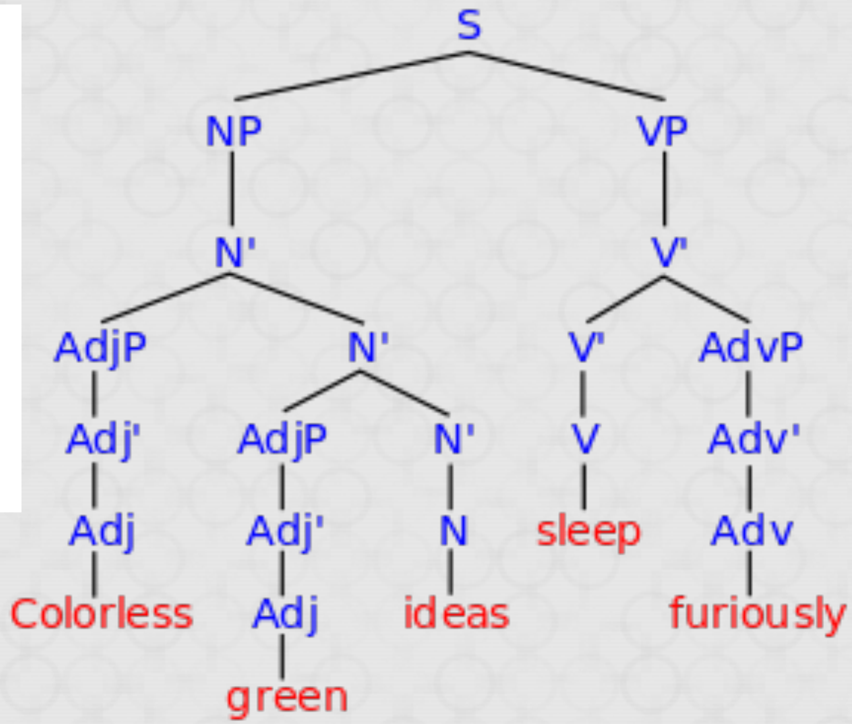
- Innate
- Poverty of the stimulus argument
- Universal Grammar
- Phrase structure tree
  - Noun Phrase
  - Verb Phrase



PHRASE TREE STRUCTURE

- Noun Phrase & Verb Phrase





- Colourless green ideas sleep furiously

STEVEN PINKER AND COLLEAGUES

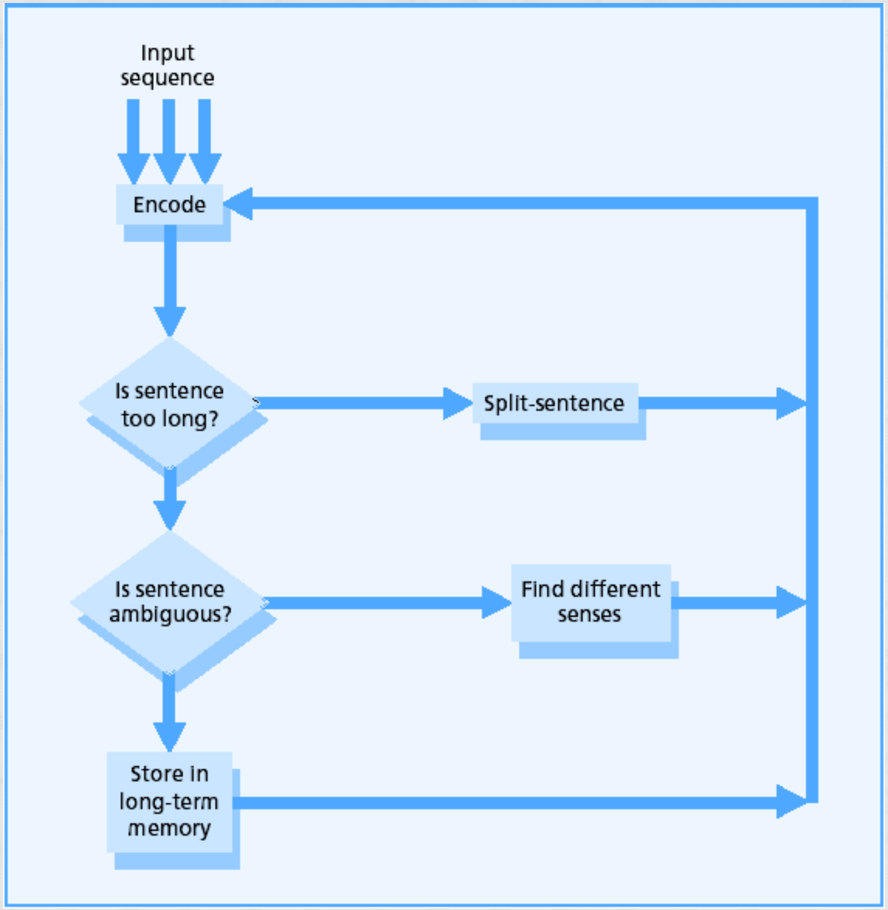
- Model of inflectional morphology (specifically English past tense)
- Abstract symbolic rules
- Two systems
  - Rule system
  - Lexical memory

PAST TENSE PROBLEM

- Regular & Irregular verbs
  - Regular - ending in -ed
    - Verb stem + past tense morpheme
    - Kick-ed, flogg-ed, patt-ed
  - Irregular - exception words (c.160)
    - Take = took not taked
    - Go = went not goed
    - buy = bought not buyed

NATIVIST WORLDVIEW

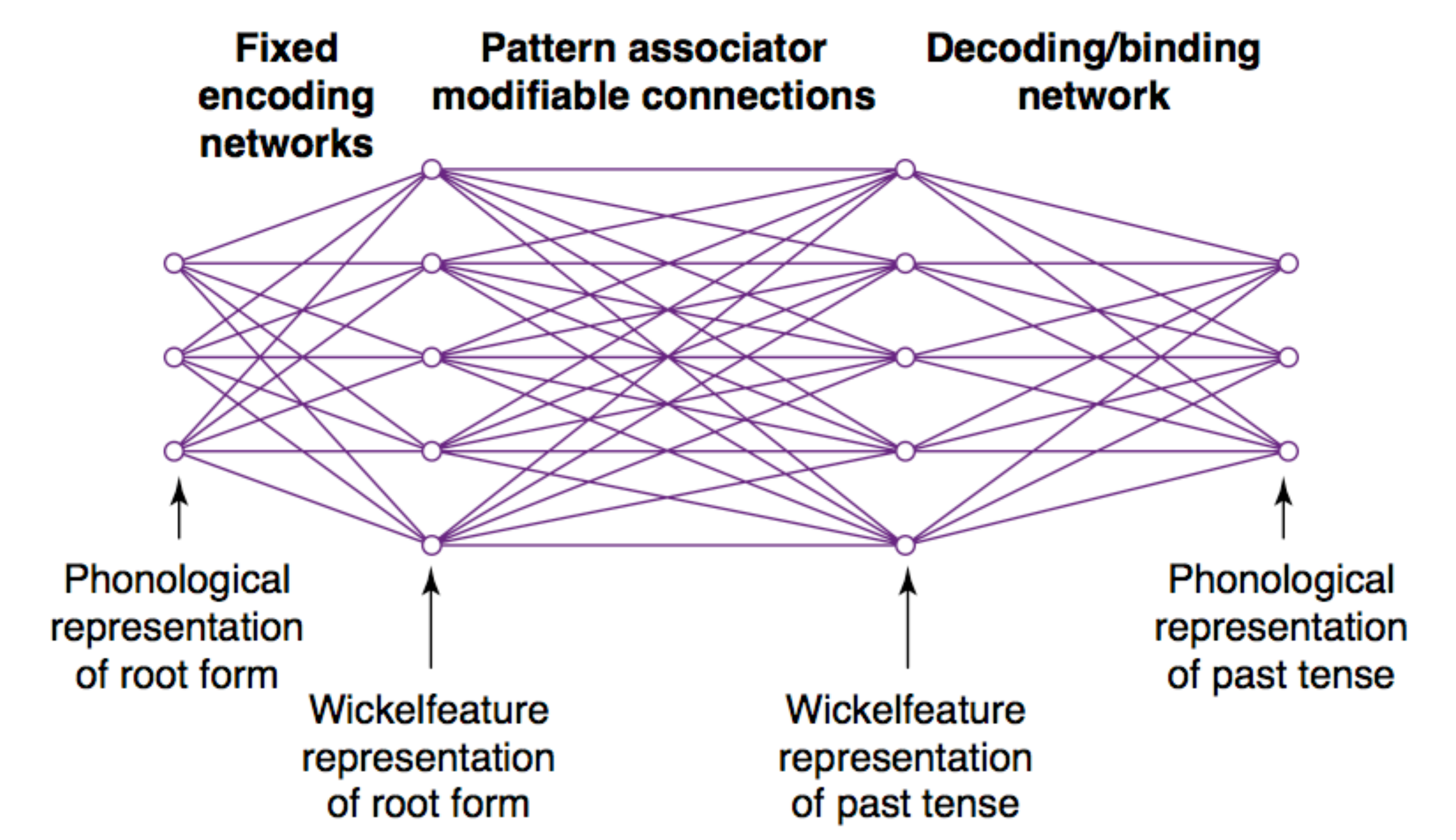
- Symbolic
  - Innate
  - Rules
  - Modular
  - Domain-specific
- Production systems
- Higher order top-down



EMPIRICIST VIEW

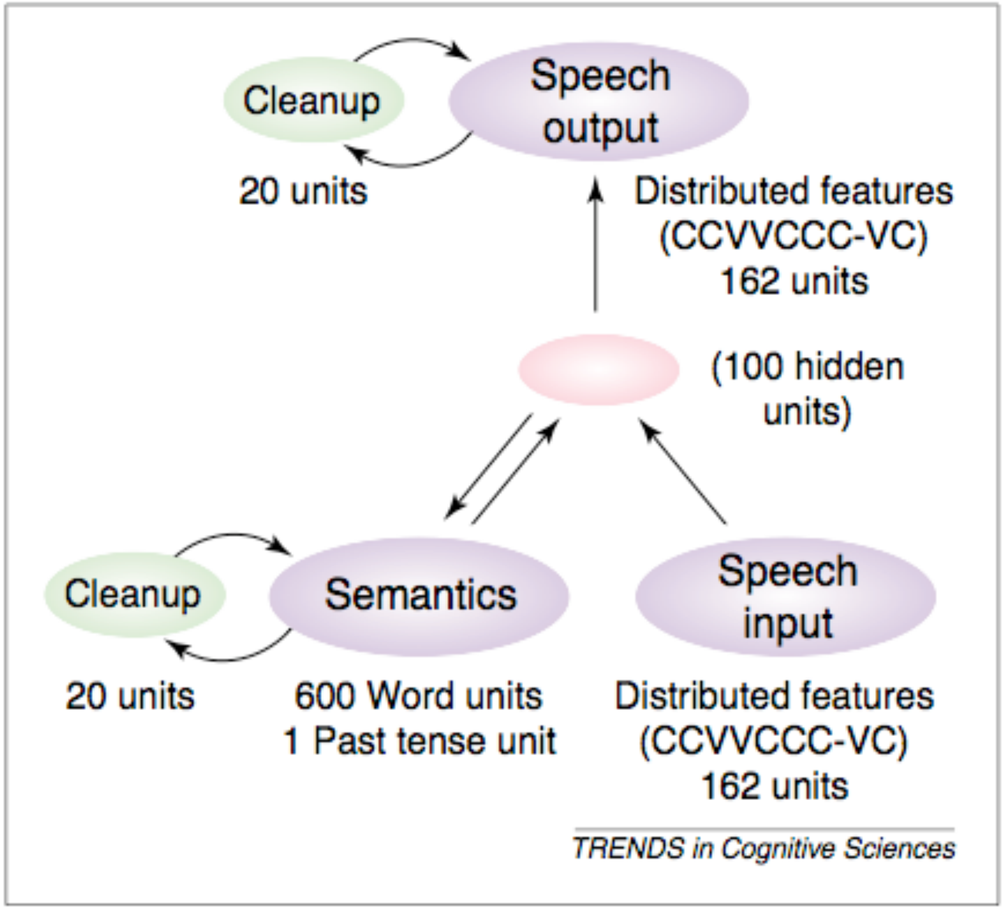
- Philosophical empiricism
- Grammatical rules are end-product
- Learning of statistical regularities
- Grammar only arises after regularities are settled
- Connectionist models of English past tense

THE RUMELHART & MCCLELLAND (1986) MODEL



JOANISSE AND SEIDENBERG (1999)

A later model that meets some of the criticisms laid against Rumelhart & McClelland model



EMPIRICIST WORLDVIEW

- Sub-symbolic
  - No explicit rules
  - Graded and Distributed
  - Statistical learning
  - Domain general
- Connectionist models
- Low-level Bottom-up approach

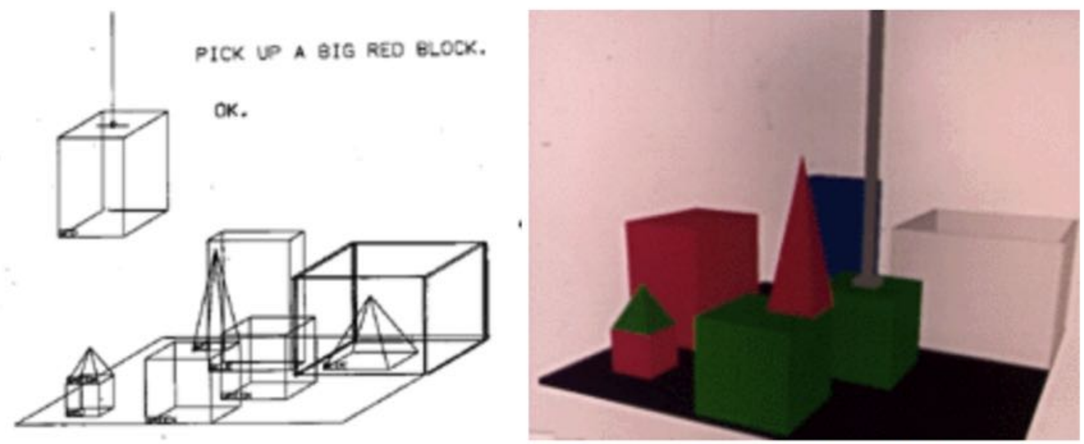
COMPUTATIONAL LINGUISTICS

The early years

THE EARLY YEARS

Winograd (1971)

- SHRDLU
  - program for understanding natural language, written by Terry Winograd at the M.I.T. Artificial Intelligence Laboratory in 1968-70



Original screen display

Later color rendering (Univ. of Utah)

THE EARLY YEARS

SHRDLU

- Person: PICK UP A BIG RED BLOCK.
- Computer: OK. (does it)
- Person: GRASP THE PYRAMID.
- Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
- Person: FIND A BLOCK WHICH IS TALLER THAN THE ONE YOU ARE HOLDING AND PUT IT INTO THE BOX.
- Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
- Computer: OK. (does it)
- Person: WHAT DOES THE BOX CONTAIN?
- Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
- Person: WHAT IS THE PYRAMID SUPPORTED BY?
- Computer: THE BOX.
- Person: HOW MANY BLOCKS ARE NOT IN THE BOX?
- Computer: FOUR OF THEM.
- Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?
- Computer: YES, THE RED CUBE.
- Person: IS IT SUPPORTED?
- Computer: YES, BY THE TABLE.
- Person: CAN THE TABLE PICK UP BLOCKS?

CLASSIC COMPUTATIONAL LINGUISTICS TWO APPROACHES

1. Classic Natural Language Processing

NATURAL LANGUAGE PROCESSING

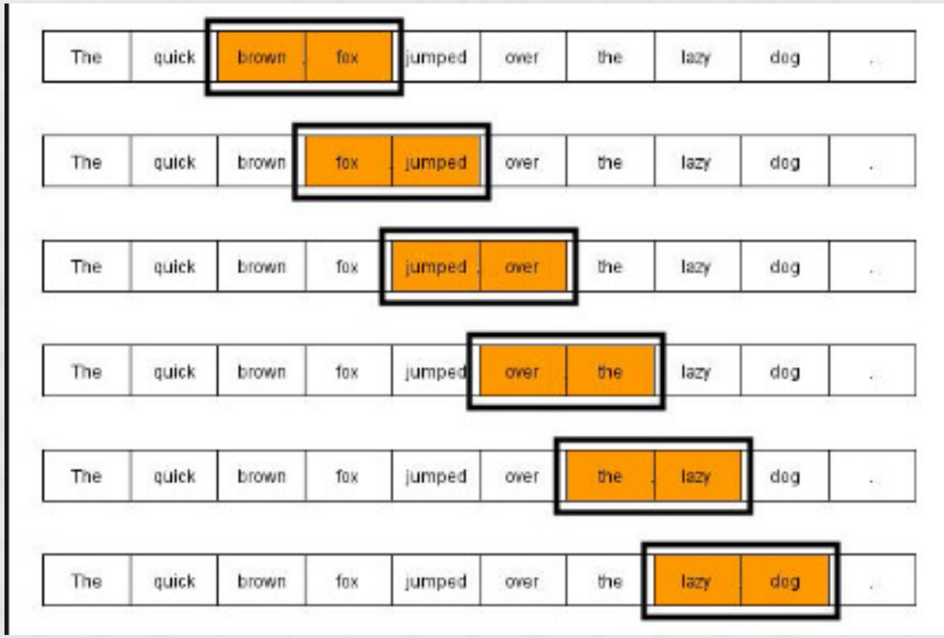
Techniques working out syntax

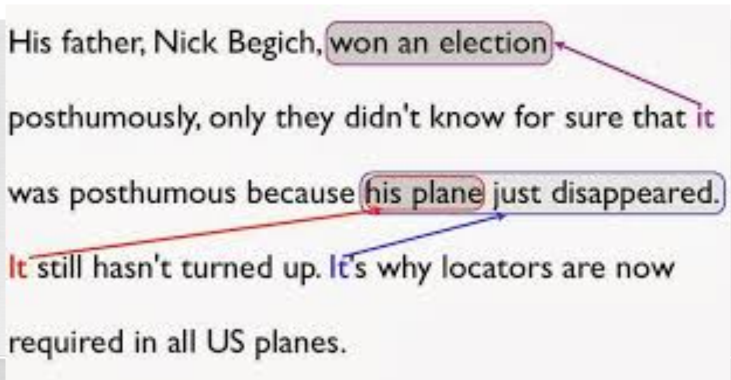
- Syntax and grammar approach to interpreting meaning
- Sentence understanding
- Syntactic parsing and tagging
- Part of speech tagging (POS)
- Anaphoric reference and disambiguation
- noun and verb phrase identifiers
- N-grams
- Tokenisation, Stemming & Lemmatisation

N-GRAM MODELS

Syntax is important

- Word order is important
- 2 -gram or bigram
- Successful for a long time
- Probability distribution over word sequences
- > about 6-grams seems to stop producing gains
- Problems of anaphoric reference




His father, Nick Begich, won an election posthumously, only they didn't know for sure that it was posthumous because his plane just disappeared. It still hasn't turned up. It's why locators are now required in all US planes.

TOKENISATION

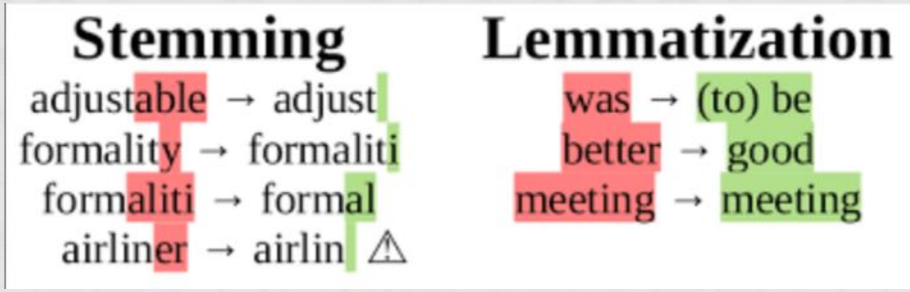
Getting basic units

- Need to decide what constitutes a token
  - A word?
  - Numbers?
  - Computer code?
  - Emojis?
- What to use? Spaces Punctuation
- Stop words
- unnest_tokens in tidytext (more bag of words)

STEMMING AND LEMMATISATION

Techniques working out syntax

- Stemming
  - Convert word to its root
  - Removes inflections prefixes, suffixes
  - based on sets of rules
- Lemmatisation
  - Converts a word to its lemma, its base form
  - based on linguistics

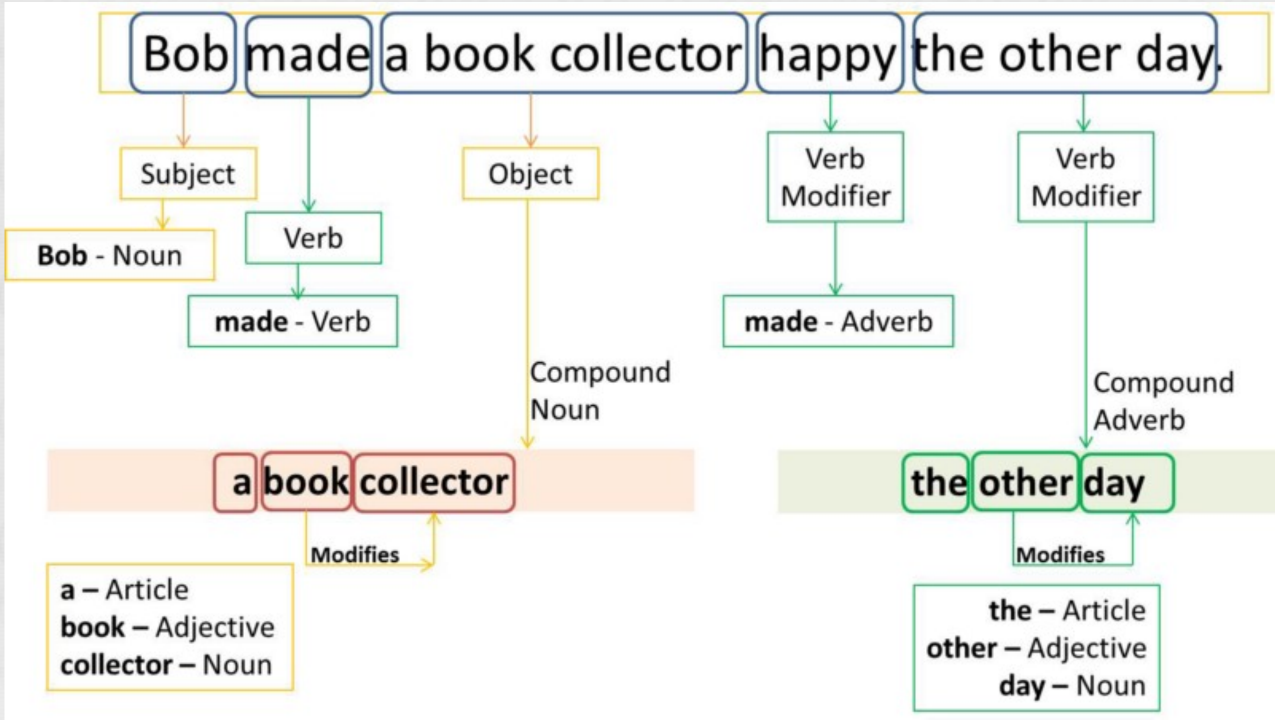


Word	*Lemmatization*	*Stemming*
was	be	wa
studies	study	studi
studying	study	study

PART OF SPEECH TAGGING

Syntax rules and grammar

- Grammatical tagging
- Labelling Verbs, Nouns, Adjectives, Adverbs
- Subject, Object of a sentence, phrase or utterance



NATURAL LANGUAGE PROCESSING

Tools

- Stanford CoreNLP
  - https://stanfordnlp.github.io/CoreNLP/
- Natural Language Toolkit - NLTK
  - http://www.nltk.org - Python based
- Apache OpenNLP
  - http://opennlp.apache.org
- GATE - General Architecture for Text Engineering
  - https://gate.ac.uk/

COMPUTATIONAL LINGUISTICS TWO APPROACHES

2. Bag of words

STATISTICAL APPROACHES

- Bag of Words
- Term Document Matrices
- Singular Value Decomposition
- Vector Space Models
- Cosine similarity
- Latent Semantic Analysis LSA, LSI (Indexing)
  - http://lsa.colorado.edu
- Probabalistic LSA
- Topic Modelling - Latent Dirichlet Allocation LDA
- Return of Terry Winograd
  - Page, L., Brin, S., Motwani, R. and Winograd, T. (1998) The PageRank Citation Ranking: Bringing Order to the Web. Technical Report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project.

STATISTICAL APPROACHES

- Pagerank
  - Page, L., Brin, S., Motwani, R. and Winograd, T. (1998) The PageRank Citation Ranking: Bringing Order to the Web. Technical Report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project.

STATISTICAL APPROACHES

- Vectorspace models

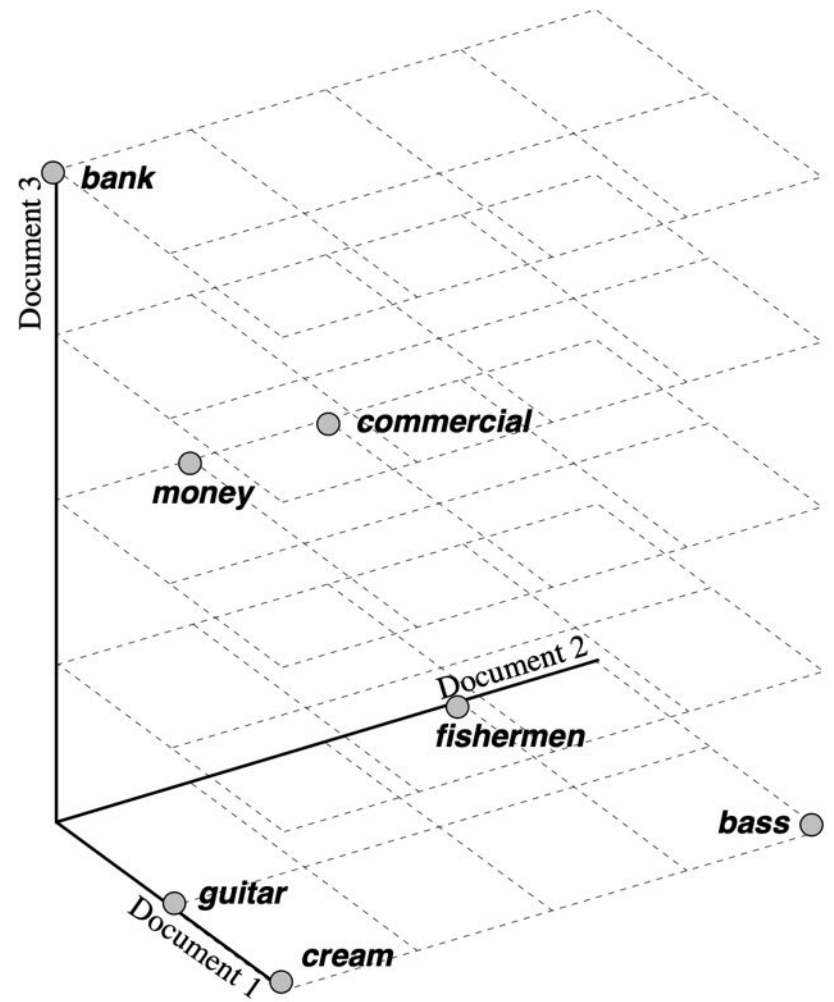
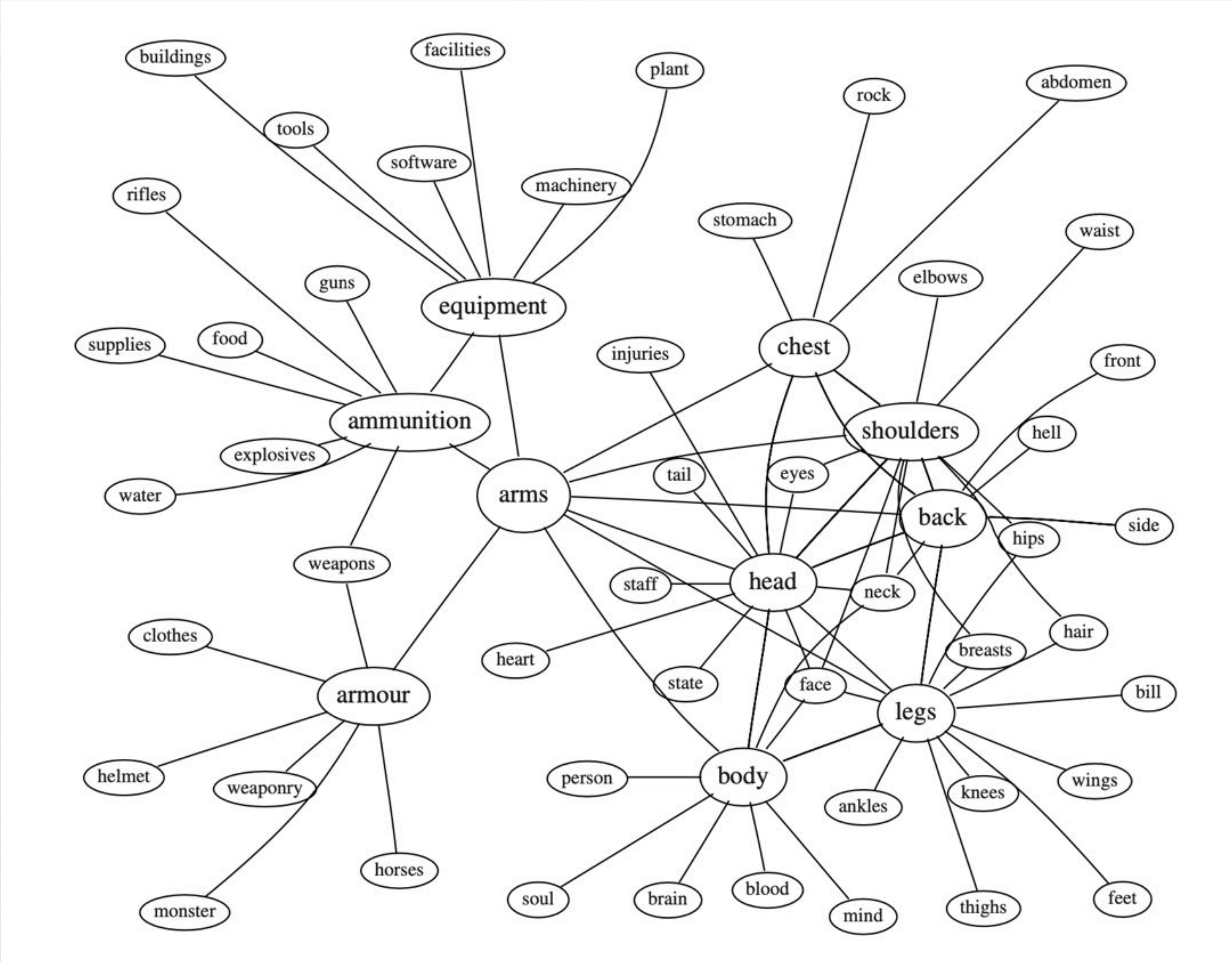


FIGURE 36 Imagining the word vectors in Table 4 as points in the three dimensional space $\mathbb{R}^3$

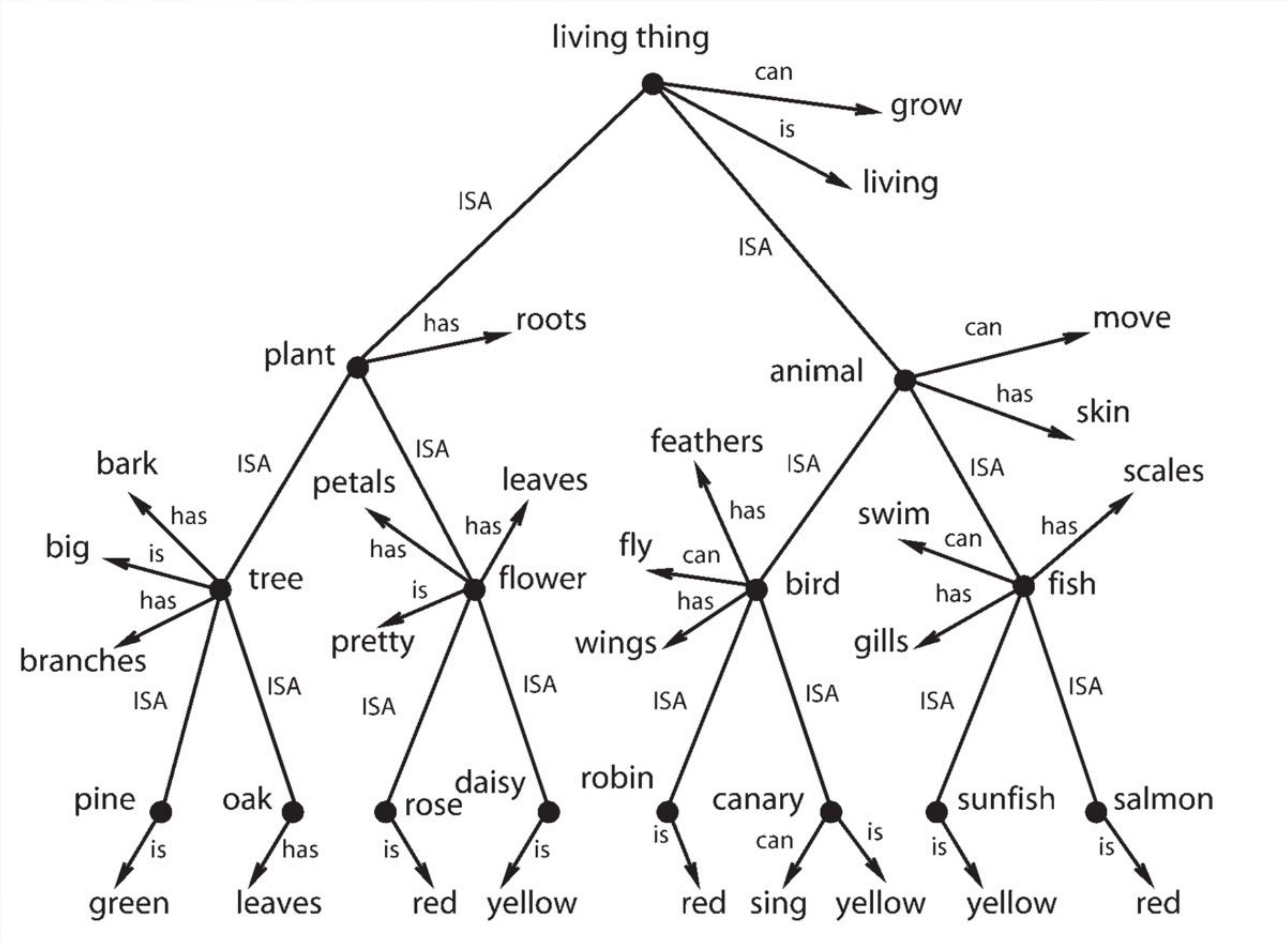
STATISTICAL APPROACHES

- Vectorspace models - graph visualisations



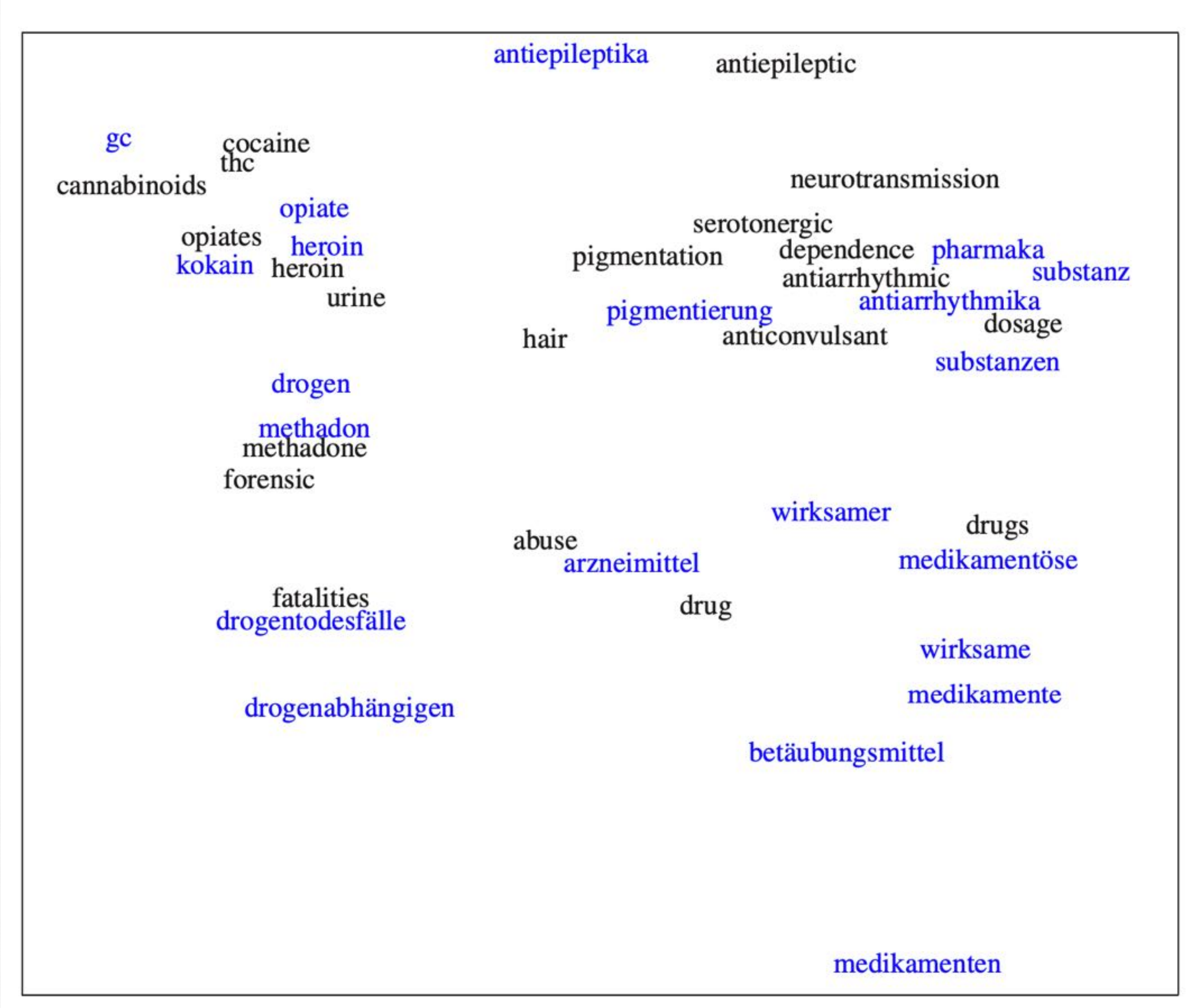
STATISTICAL APPROACHES

- Collins & Quillion (1969) human memory of knowledge



STATISTICAL APPROACHES

- Bilingual corpora and models



STATISTICAL APPROACHES

- Semantic Networks have primacy over syntax
  - Wordnet - lexical database wordnet.princeton.edu
  - Wordnet-Affect
- Open Mind Common Sense
  - ConceptNet
  - from MIT - Marvin Minsky, Push Singh, Catherine Havasi, Rob Speer
    - https://luminoso.com
- SenticNet and Sentic Computing
  - Erik Cambria

STATISTICAL APPROACHES

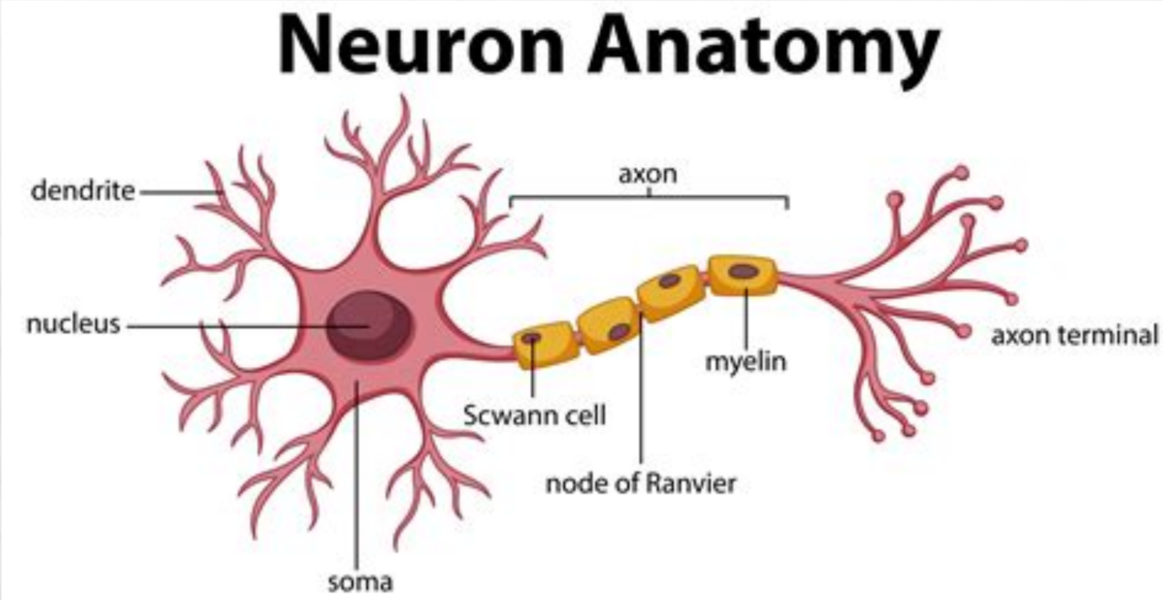
NEURAL NATURAL LANGUAGE PROCESSING

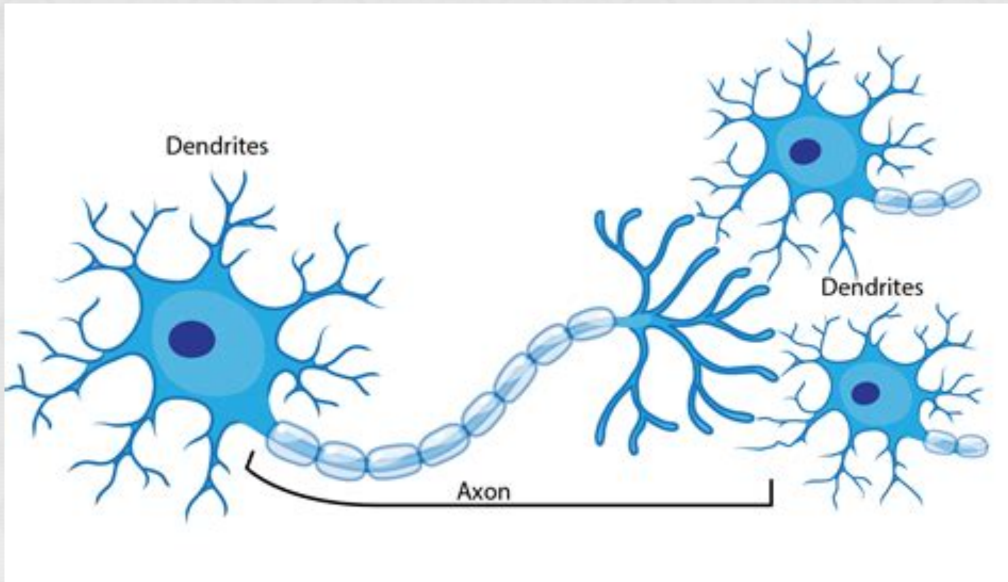
2010 - Present

NEURONS

2013

- Axon
- Dendrite
- Nucleus
- Synapses
- Myelin
- Electro-chemical signals
- Neurotransmitters
  - Weights
- Sigmoid like activation

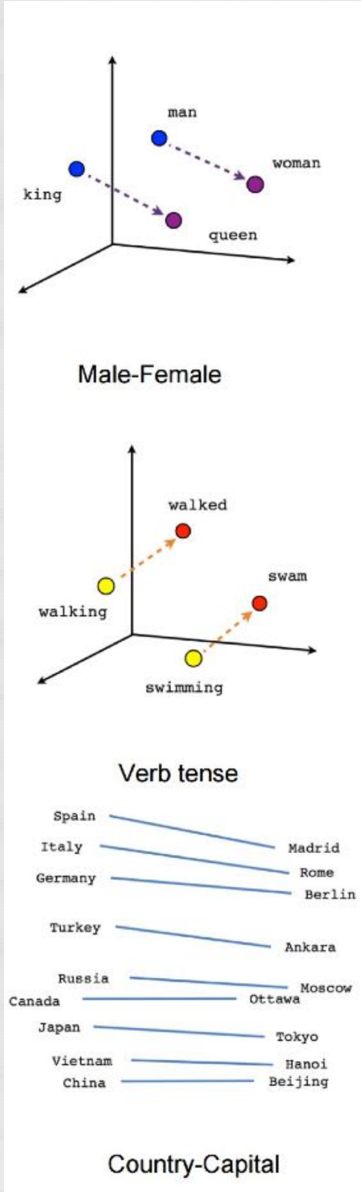




WORD2VEC

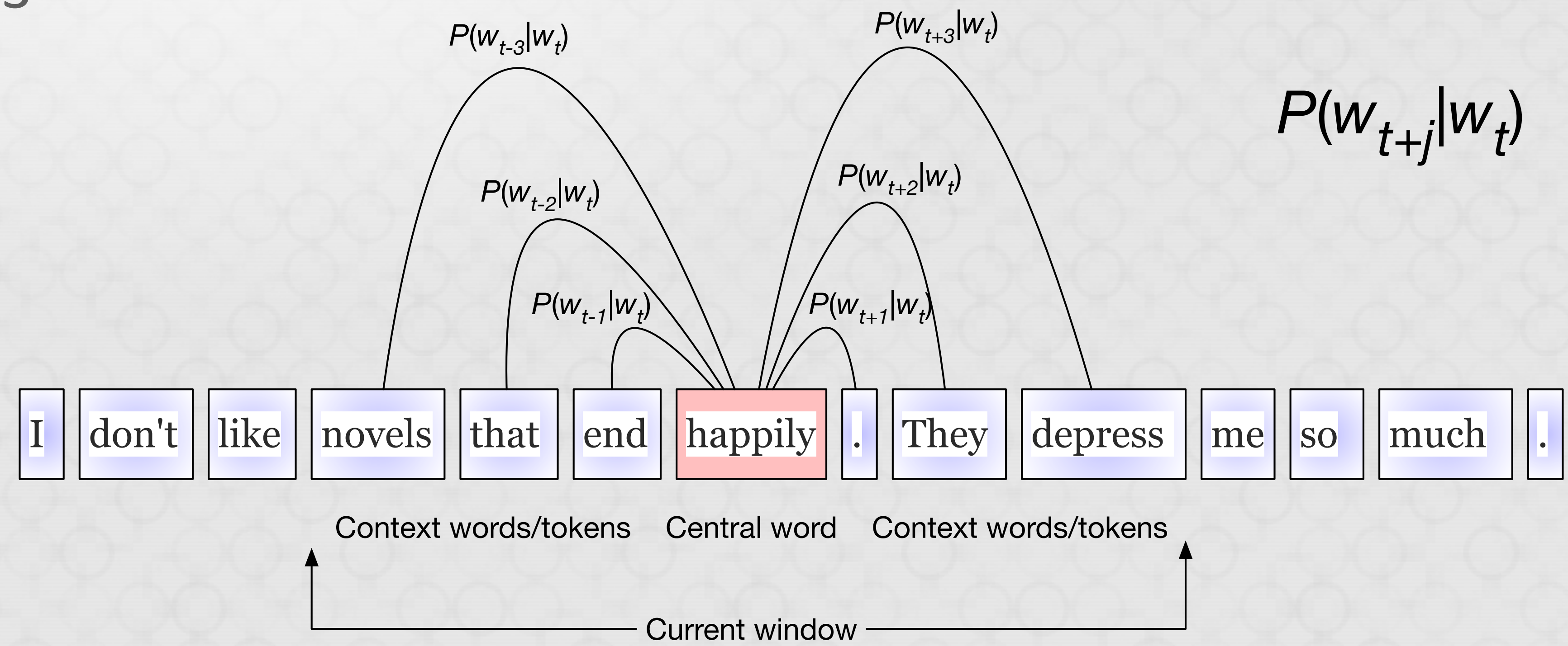
2013

- Both word vector and SVD methods are computationally very expensive.
- Vectors themselves are great
- Word2Vec neural network capturing both semantic and syntactic information
- Sum of a lot of windows of text that capture context of words



WORD2VEC

2013

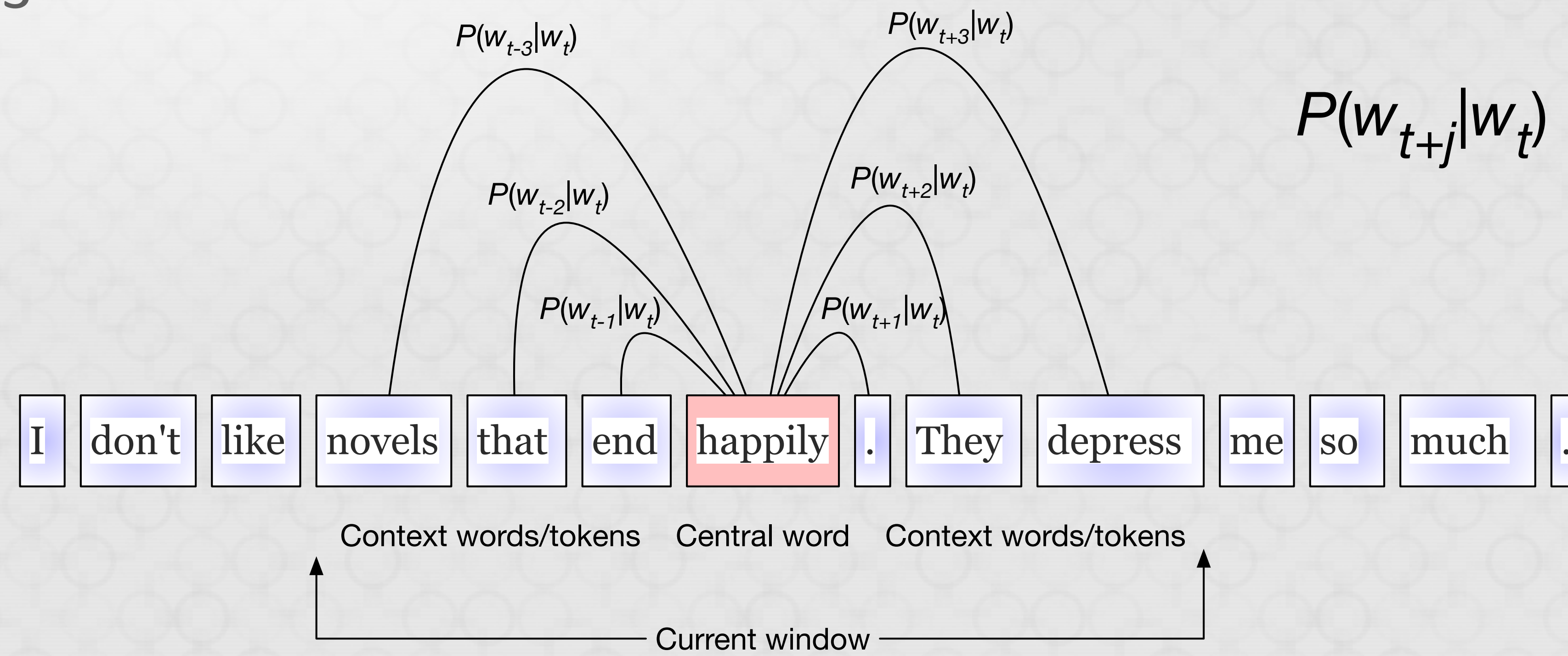


$P(w_{t+j}|w_t)$

- Lots of these windows are calculated for a large amount of text to get a probability distribution
- 2 vectors per word – one for word in centre another for word as context

WORD2VEC

2013



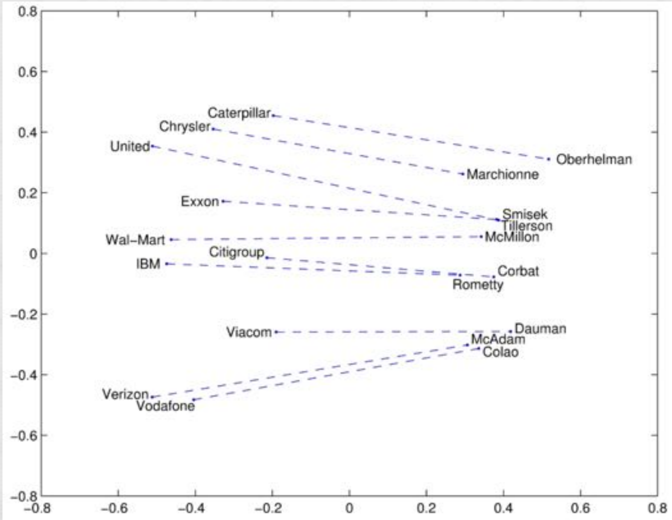
- Skip-gram models as above - predict context words given centre word
- Continuous bag-of-words (CBOW) – predict metre word from context bag of words

GLOVE

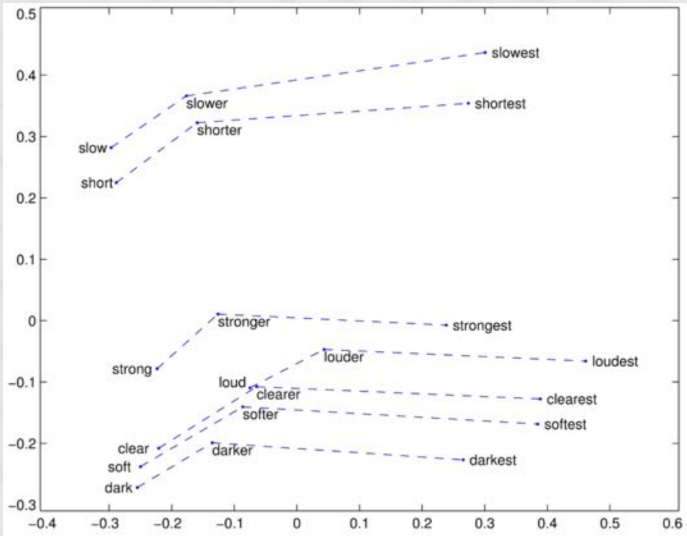
2014

- GloVe: Global Vectors for Word Representation
- A word vector space is created
- The neural network is trained on the word co-occurrence statistics.
- Best of both worlds
    - co-occurrence and neural
- Good for finding relationships

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).



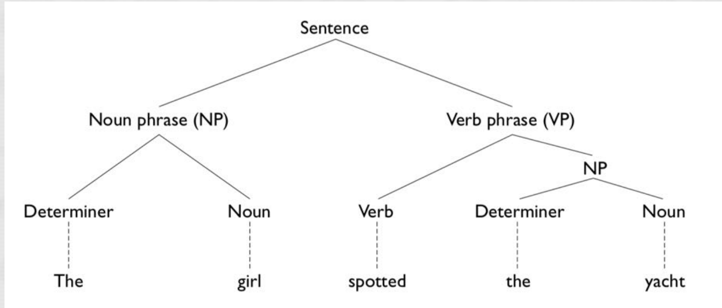
Company - CEO

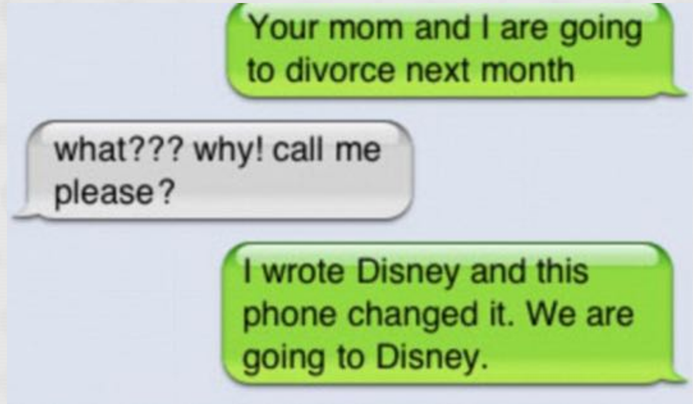


Comparative - Superlative

LINGUISTIC TASKS

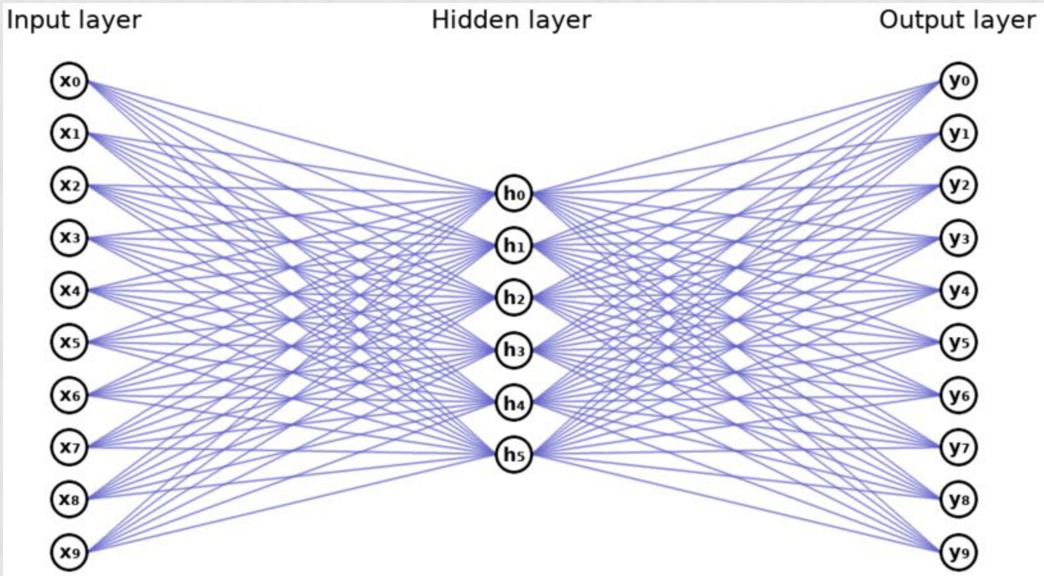
- Dependency parsing - a bit like Chomsky's Generative grammars
- Language models - predictive text, autocorrect - basis of a lot of tasks
- Machine translation
- Speech Recognition
- Grammar Correction
- Summarisation
- Dialogue modelling
- Handwriting recognition
- Sentiment Analysis

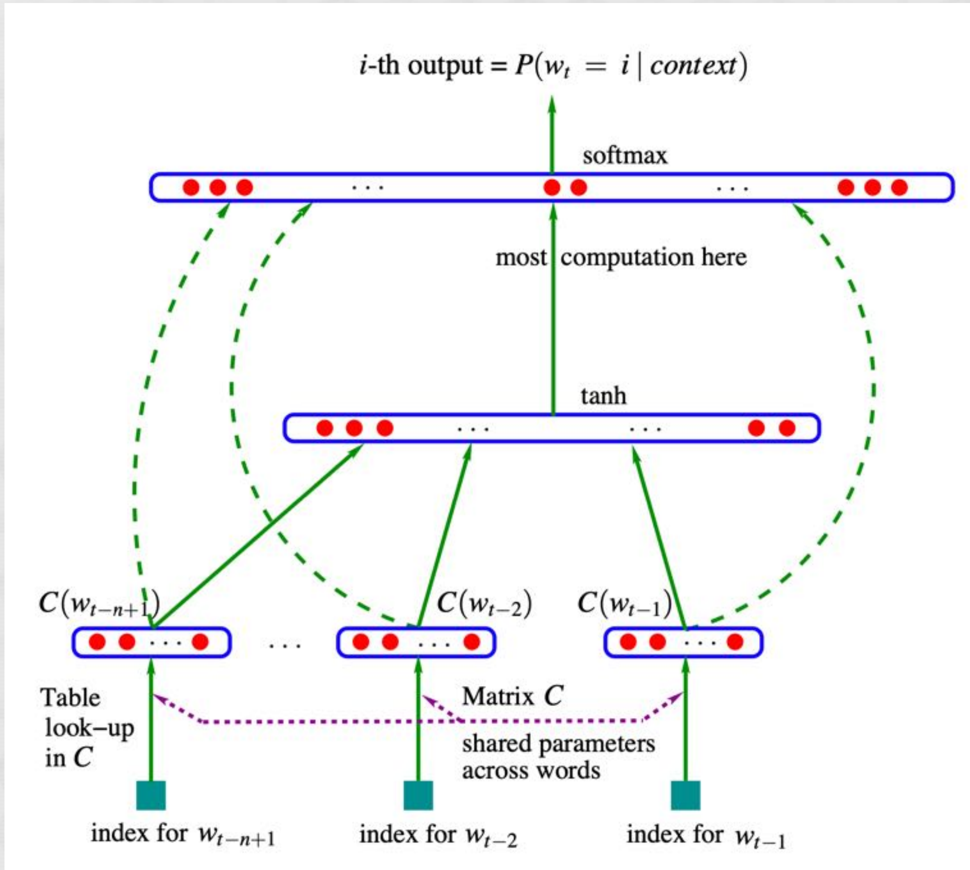




BASIC NEURAL MODELS

- Distributed representations





Bengio, Y., Ducharme, R., & Vincent, P. (2000). A neural probabilistic language model. *Advances in Neural Information Processing Systems*, *13*.

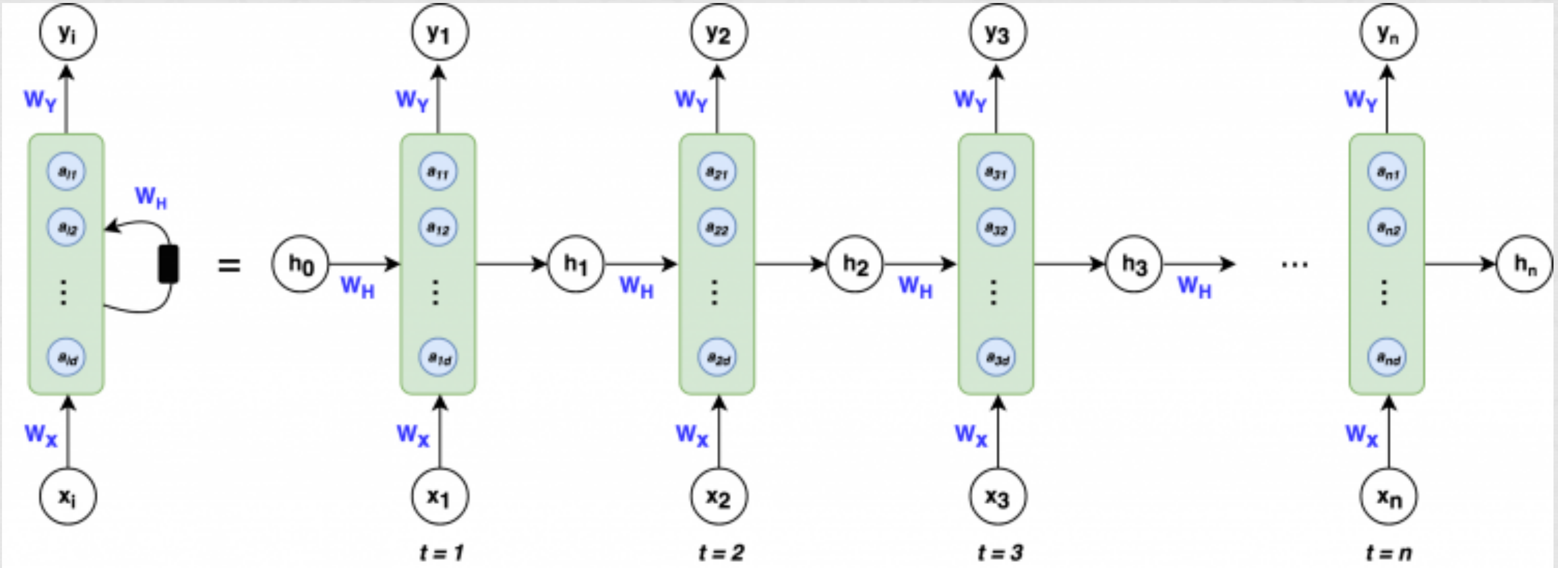
RNN

- RNN - Recurrent Neural Networks
- Hidden layer gets maintained over time and fed back into itself

y_i = first predicted next word output vector

Hidden layer

x_i = first word input vector



- Can retain information from previous steps - context - in theory at any rate (vanishing gradients).

LONG SHORT TERM MEMORY RNNS

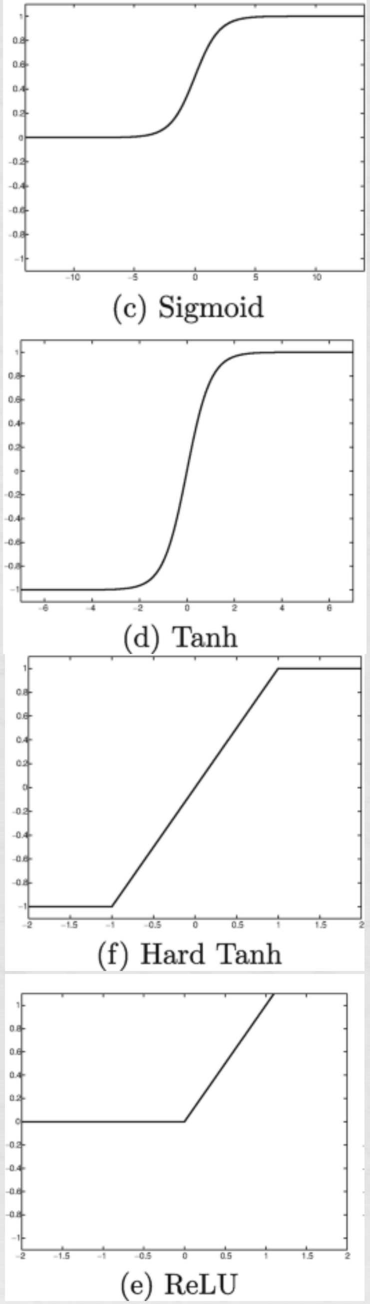
1997/2000

- LSTM - Long Short Term Memory
- Solution to vanishing gradients
- Vanishing gradients mean hidden states are overwritten
- Add another hidden layer (cell state) to act as a long term memory (they just store a vector)
- Analogy with computer RAM - read, write, and erase things in the cell
- Storage decisions (read, write, and erase) are made by gates - output gate, input gate, forget gate
- Gates have values between 0 and 1

NON-LINEAR FUNCTIONS

Activation Functions

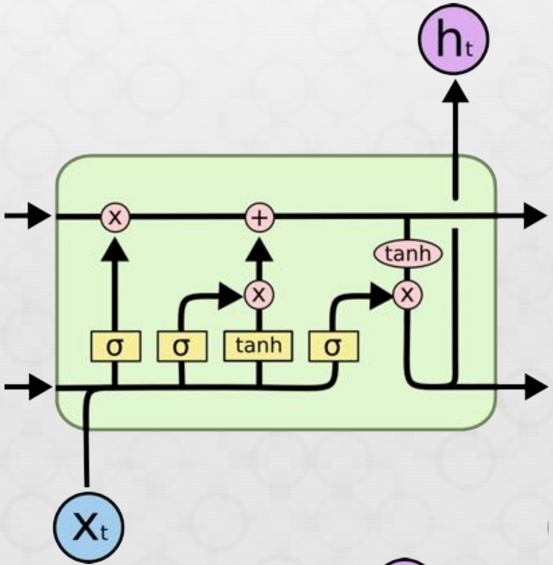
- Sigmoid - biologically realistic 0 to 1
- tanh - hyperbolic tan similar to sigmoid but -1 to 1
- Hard tanh - simplified tanh, flat and linear
- ReLU - Rectified Linear Unit
  - Simplification of most important aspects - extends beyond 1.

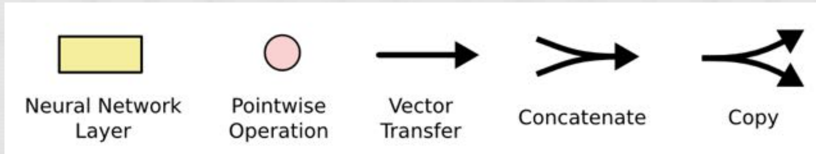


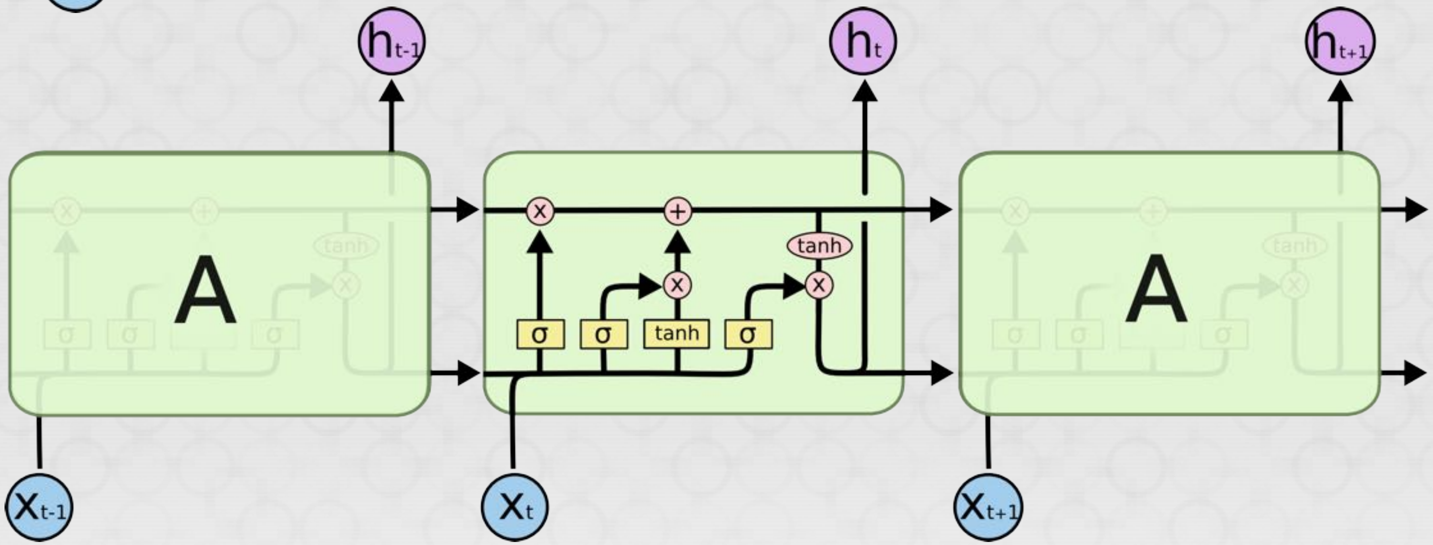
https://link.springer.com/chapter/10.1007/978-3-319-94463-0_1

LONG SHORT TERM MEMORY RNNS

First in 1997/2000 - picked up later







https://colah.github.io/posts/2015-08-Understanding-LSTMs/

TRANSFORMERS

Attention

- Attention
- Adds another hidden vector with a distribution of attention - highlights more important information
- Encoder and decoder parts - from machine translation
- Allows further looking back at crucial aspects than LSTM

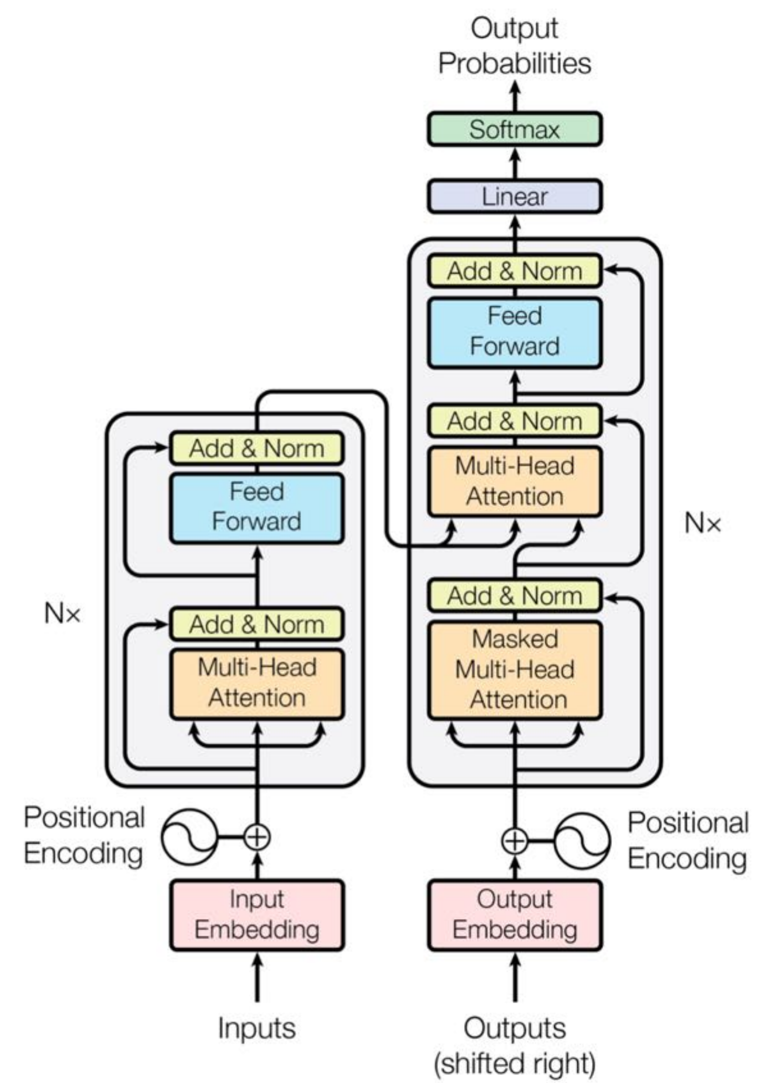


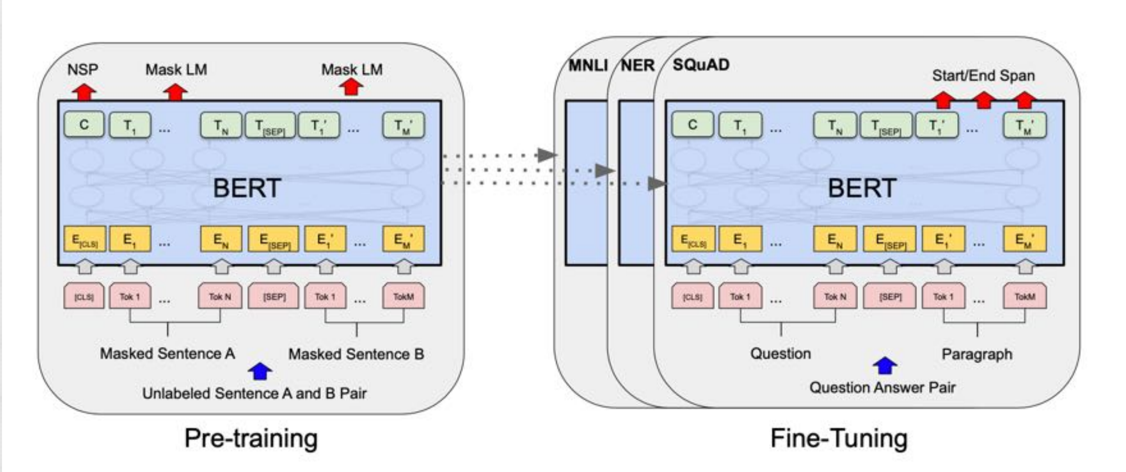
Figure 1: The Transformer - model architecture.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

TRANSFORMERS

BERT

- Bidirectional Encoder Representations from Transformers
- Two step process pre-training and fine-tuning
- Multiple attention heads
- Led to many descendants
- See Hugging Face and spaCy



Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

TRANSFORMERS

BERT

- Pre-trained models
- BERT - Google
- Roberta
- ELMo (Em-beddings from Language Models)
- Megatron-LM - Nvidia, Microsoft
- OPEN AI GPT-2 (Generative Pre-Training)
- OPEN AI GPT-3 - Current champion (have to ask for access)
    - Elon Musk

TRANSFORMERS

Megatron-LM 2019

- Megatron-LM - Nvidia, Microsoft
- Largest ever model
- Training Billion+ Parameter Language Models Using GPU Model Parallelism
- 24x the size of BERT and 5.6x the size of GPT-2
- The model is trained on 174GB of text, requiring 12 ZettaFLOPs over 9.2 days to converge
- Bigger is not necessarily better

SENTIMENT ANALYSIS

SENTIMENT ANALYSIS

- Dual Processing approaches
- Subjective Utility and normative theories - not working
- Kahnemann and Tversky (1973)
- Emotion had been ignored
- Emotion is important.
- Dual Processing - thinking fast and slow
- Need to understand emotion in communication

SENTIMENT ANALYSIS

- Bag of words and statistical approaches
- Semantic networks?
  - Where does emotion fit in?
  - Emotion is messy?
  - Emotion is not symbolic?
  - but some symbols relate to emotion
    - words like Ekman's big six: anger, happy, sad, surprise, fear, disgust.
    - and many more

SENTIMENT ANALYSIS

- Word counting approaches
  - Create a list of words that relate to emotion. Score based on list
  - Positive emotion list vs Negative emotion list.
  - Opinion mining - polarity
  - What about:
    - "I want a burrito so bad"
    - "I just had a burrito. It was so bad."
    - Sarcasm
    - negation - "not bad"
  - Bag of words and vectorspace suffer here
    - probably okay at larger statistical levels

SENTIMENT ANALYSIS

Linguistic Inquiry and Word Count (LIWC)

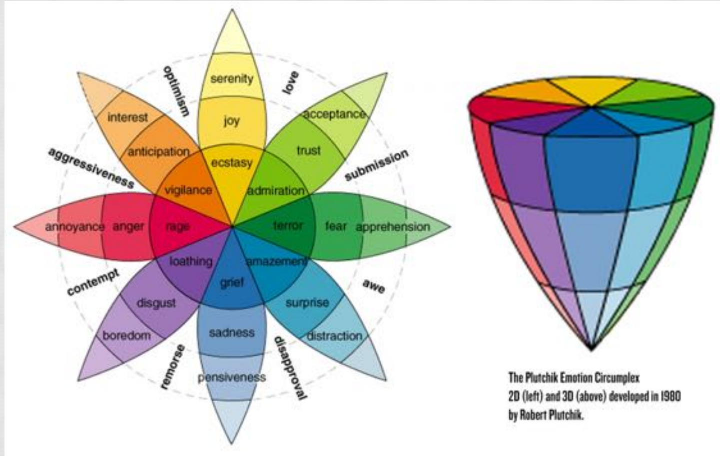
- Dictionary with norms - word count
- Psychology validation approach rather than a machine learning "ground truth"
- Years of collecting norms about words and their meanings.
- Made into a dictionary
- 100,000 files of texts - 250 million words
- LIWC-22, LIWC 2015, LIWC 2007, LIWC 2001. (www.liwc.app)
- Many scores including affect/emotions

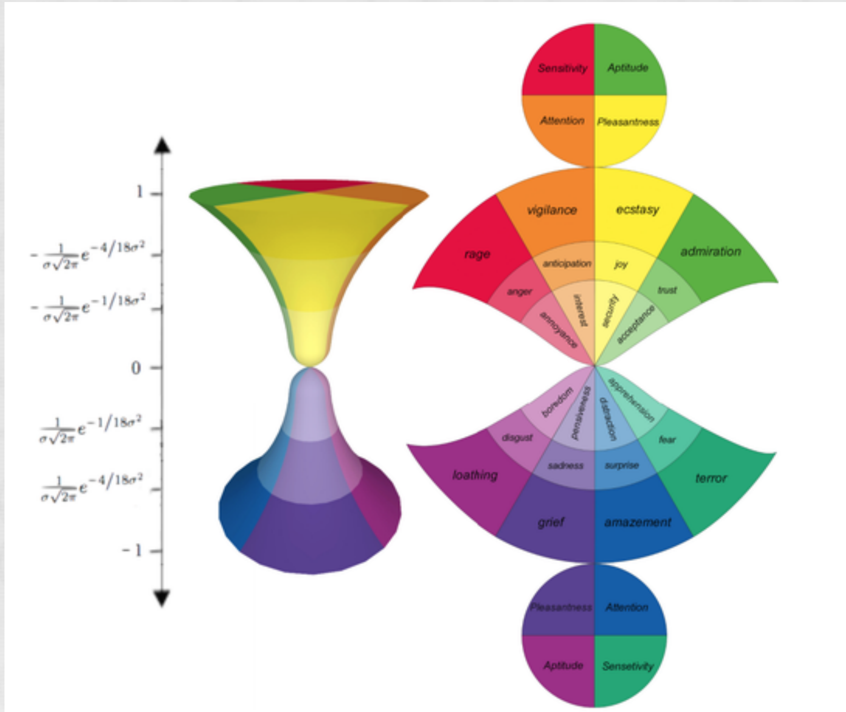
SENTIMENT ANALYSIS

- Additions of NLP and Bag of words approaches to minimise problems
- Many words are new or proper nouns especially brands
    - Google, Apple, Microsoft, Facebook
    - not in the dictionary
- Common Language and common sense required
- Polysemy, disambiguation, context
- Pragmatics - Relevance Theory

SENTIC COMPUTING

- Erik Cambria
  - Coming out of MIT
    - Luminoso group
  - AffectNet and ConceptNet
  - Common Sense
  - Semantic Space blending
    - combining vectorspaces
  - Plutchik's wheel of emotion model?
    - hourglass model?
  - 4 dimensions - 24 labels





SENTIMENT ANALYSIS IN R

- Datacamp
  - 3 Assignments
    - Introduction to Text Analysis in R
    - Text Mining with Bag-of-Words in R
    - Sentiment Analysis in R
- Packages
  - SentimentAnalysis
  - sentimentr
  - tidytext
  - textir

See also: https://CRAN.R-project.org/view=NaturalLanguageProcessing

Also many Python packages and courses - probably more

SENTIMENT ANALYSIS

Important

- Emotion requires an autonomic nervous system
- Emotion is in the person - not in the word or in the signal.
- The signal or word is evidence for the meaning but does not contain meaning in itself

- *Brandwatch - SA Brand Analytics company*
  "The insight that can be gained from large datasets (millions of Tweets) will overshadow the concerns about reliability at a granular level (a single Tweet)".

EMOJIS

- :) :p 😀😃😄😁😆😅😂🙂🤣☺️😊😇🙂😍🙃😉
- Surely about emotion
  - Yes but not felt emotion
- Pictographic, symbolic, and discrete
- lol
- gifs
  - More continuous - still a bit symbolic
- Innovative quotatives - “then I was like…”
- Disambiguation, adding more evidence of meaning

MODERN NLP AND SENTIMENT ANALYSIS

- BERT approaches to classification of negative and positive
- Dependent on the training sets.
- Knowledge-Enriched Transformer for Emotion Detection in Textual Conversations Zhong et al.
- Combines Transformer with scores of NRC VAD words (saifmohammad.com/WebPages/lexicons.html https://vimeo.com/285800613)
- BERTweet - Nguyen (2020)
- Humor Knowledge Enriched Transformer (Hasan et al. (2021)
- Emotional RobBERT (Dutch) De Bruyne et al (2021)

PARALANGUAGE

PARALANGUAGE

and other phenomena

- Prosody
- Pitch
- Volume - intensity
- Intonation
- "Yeah, yeah."
- Backchanneling
- Un huh, hmmm, ugh
- Nods, head shakes, gestural deixis
- Linguistic acronyms
  - OMG, WTF, IMHO, AFAIK, IIRC, LMAO
- Swear words - Taboo language

HOW DO WE SPEAK?

Transcription again

THE CODING/CLASSIFICATION SYSTEM

Laughter - Conversation Analysis (CA)

(3)[Holt:2:2:5–6]

1	Lesley:	I hope you don't mind your conversations being re↓corded
2		this telephone is bugged.
3		(0.2)
4	Bond:	↑Oh ↓is it?=
5		(0.2)
6	Bond:	↑ha [ha ha ha[ha h[a ha ha ha
7	Lesley:	[Well- [K- [Kathrine's doing 'er thesis o:r or
8		something on um
9	Bond:	Oh[that's right you told me she wz going to,
10	Lesley:	[speech.
11	Lesley:	Yes.
12	Bond:	Ye:s.
13		(0.2)
14	Bond:	Oh we:ll
15	Bond:	Down it's down f'posterity [hey,
16	Lesley:	[No STATE secrets,=
17	Bond: →	=[ha]ha haha eh oh]
18	Lesley:→	=[hh]heh heh heh.hh] .hh Okay th[en,]
19	Bond:	[Y e]s .hh Well-
20:	Bond:	ekhh-hu Le:mme know if you don't get that ballot
21		paper'n I'll [(check[ )]
22	Lesley:	[Yes. [O]kay

Gail Jefferson
1938-2008

Holt, E. (2010). The last laugh: Shared laughter and topic termination. Journal of Pragmatics, 42(6), 1513–1525.

CONCLUSIONS & DISCUSSION