Első óra
Ismerkedés, keretek, mi az NLP, az NLP története

Makrai Márton

Számítógépes nyelvészet @ tlp 2022 tavasz

Bemutatkozás, keretek

- az előadó Makrai Márton
  - nem Sass Bálint és Mittelholcz Iván
  - BME matematikus, MSc
  - SZTAKI, NYI, ELKH TTK, Szegedi Egyetem
  - NYTK–ELTE elméleti nyelvészeten PhD hallgató friss szigorlattal
- a hallgatók
  - név, tlp-n BA vagy MA, évfolyam, másik szak, téma, motiváció
- keretek: időpont; zoom majd valószínűleg többnyire jelenlét
- sokszor meghívott előadó lesz
- felvétel
- minden héten házi feladat
- kapcsolat
  - levlista groups.google.com/group/nlp22
  - ímél makrai.hlt@gmail.com
  - honlap https://hlt.bme.hu/hu/course/nlp22
- tankönyv: https://web.stanford.edu/~jurafsky/slp3/

Számítógépes nyelvészet és nyelvtechnológia

- számítógépes nyelvészet *(computational linguistics)*
- nyelvtechnológia (*natural language processing*, NLP)

Számítógépes nyelvészet és nyelvtechnológia

- számítógépes nyelvészet *(computational linguistics)*
  - a nyelv megismerése számítógépes modellek segítségével
- nyelvtechnológia (*natural language processing*, NLP)

Számítógépes nyelvészet és nyelvtechnológia

- számítógépes nyelvészet *(computational linguistics)*
  - a nyelv megismerése számítógépes modellek segítségével
- nyelvtechnológia (*natural language processing*, NLP)
- (szabad, strukturálatlan) szövegek (hatékony) feldolgozása
  beszéd generálása
- az NLP további nevei
  - gépi szövegértés, természetesnyelv-feldolgozás
  - natural human language understanding technologies

Feladatokra példák

- Dave Bowman: Open the pod bay doors, HAL.
  HAL: I'm sorry Dave, I'm afraid I can't do that.
  - Stanley Kubrick and Arthur C. Clarke, screenplay of 2001: A Space Odyssey
- gépi fordítás
- webalapú kérdésmegválaszolás
  - What does "divergent" mean?
  - What year was Abraham Lincoln born?
  - How many states were in the United States that year?
  - How much Chinese silk was exported to England by the end of the 18th century?
  - What do scientists think about the ethics of human cloning?

A három nyúl

Web-based question answering

- definition questions, or simple factoid questions
  - like dates and locations
- more complicated questions
  - extracting information that is embedded in other text on a Web page
- inference (drawing conclusions based on known facts)
- synthesizing and summarizing information from multiple sources
- components include
  - information extraction, word sense disambiguation

Mit kell hozzá tudni a nyelvről/beszédről?

Nyelvi szintek

- eg the Unix `wc` program
  - counts the total number of bytes, words, and lines
- speech recognition and speech synthesis: phonetics and phonology
- Producing and recognizing word forms: morphology
- sentence structure, word order: syntax
- lexical semantics, the meaning of all the words (*export* or *silk*)
- compositional semantics
  - what exactly constitutes Western Europe
    as opposed to Eastern or Southern Europe
  - what does *end* mean when combined with *the 18th century*
- relationship of the words to the syntactic structure
  - eg *by* prase goal or agt
- Dave's utterance is a request for action
  - as opposed to a simple statement
- be polite to Dave: pragmatic or dialogue knowledge
- coreference resolution: how words like *that* or pronouns refer to previous referents

Többértelműség

- *Lementem a boltba*
- *I made her duck*
  - I cooked waterfowl for her
  - I cooked waterfowl belonging to her
  - I created the (plaster [gipsz]?) duck she owns
  - I caused her to quickly lower her head or body
  - I waved my magic wand and turned her into undifferentiated waterfowl
- *duck* and *her*
  - morphologically or syntactically ambiguous in their POS
  - *duck* can be a verb or a noun; *her* can be a dative pronoun or a possesive
- *make* is
  - semantically ambiguous; it can mean create or cook. Finally, the verb
  - syntactically ambiguous in a different way
    - transitive, ditransitive, or take a direct object and a verb
- in a spoken sentence,
  the first word could have been eye or the 2nd maid

Szövegfeldolgozás mint a többértelműség feloldása

- lexical disambiguation
  - deciding whether *duck* is a verb or a noun can be solved by POS tagging
  - whether *make* means “create” or “cook” can be solved by word sense disambig
  - text-to-speech synthesis system reading the word *lead* needs to decide whether it should be pronounced as in *lead pipe* or as in *lead me on*
- syntactic disambiguation
  - whether *her* and *duck* are part of the same entity
  - can be addressed by probabilistic parsing
  - whether a sentence is a statement or a question: speech act interpret

Áttekintés

Turing test

- if computers can process language as humans do: intelligent machines
- language is intertwined with our general cognitive abilities
- Alan Turing (1950)
- three participants: two people and a computer
- the interrogator must determine which of the other two participants is human

I am unhappy.
DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY
I need some help, that much seems certain.
WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP
Perhaps I could learn to get along with my mother.
TELL ME MORE ABOUT YOUR FAMILY
My mother takes care of me.
WHO ELSE IN YOU FAMILY TAKES CARE OF YOU
My father.
YOUR FATHER
You are like my father in some ways.

ELIZA

- Weizenbaum (1966)
- imitating the responses of a Rogerian psychotherapist
- pattern matching
  - "I need X" $\rightarrow$ "What would it mean to you if you got X?"
- no need to know anything to mimic a Rogerian psychotherapist
  - one of the few dialogue genres where listeners can act as if they know nothing of the world
- many people who interacted with ELIZA came to believe that it really understood them and their problems,
- many continued to believe in ELIZA's abilities even after the program's operation was explained to them
- modern conversational agents
  - answer questions, book flights, or find restaurants
- regular expressions

Társunk, a gép

- social sciences confirmed another of Turing's predictions
- at the end of the century the use of words and educated opinion will have altered so much that we will be able to speak of machines thinking without expecting to be contradicted
- regardless of what people believe or know about their inner workings
- people talk about them and interact with them as social entities
- people are polite to them, treat them as team members
  - expect that computers should be able to understand their needs and be capable of interacting with them naturally
- when a computer asked a human to evaluate how well the computer had been doing, the human gives more positive responses than when a different computer asks the same questions (Reeves and Nass 1996)
- people also give computers higher performance ratings if the computer has recently said something flattering to the human
- speech- and language-based systems may provide many users with the most natural interface for many applications

Áttekintés

Automaták, valószínűség, információ-elmélet I

1940-es és 50-es évek

- Turing's (1936) model of algorithmic computation
  - the foundation of modern computer science
- McCulloch-Pitts neuron (McCulloch and Pitts, 1943)
  - a simplified model of the neuron as a kind of computing element that could be described in terms of propositional logic
- Kleene (1951, 1956) on finite automata and regular expressions
- Shannon (1948)
  - probabilistic models of discrete Markov processes applied to automata for language
- Chomsky (1956): finite-state machines as a way to characterize a grammar
  - finite-state language as a language generated by a finite-state grammar
- formal language theory used algebra and set theory

Automaták, valószínűség, információ-elmélet II

1940-es és 50-es évek

- context-free grammar, first defined by Chomsky (1956) for natural langs
  - independently discovered by Backus (1959) and Naur+ (1960) in their descriptions of the ALGOL programming language

Automaták, valószínűség, információ-elmélet III

1940-es és 50-es évek

- probabilistic algorithms for speech and language processing
- Shannon: the metaphor of the noisy channel and decoding
  - for the transmission of language through media such as communication channels and speech acoustics
- the concept of entropy from thermodynamics
  - measures the information capacity of a channel, or the information content of a language
- the first measure of the entropy of English
  - by using probabilistic techniques
- sound spectrograph (Koenig+ 1946)
- instrumental phonetics
  - groundwork for later work in speech recognition
- machine speech recognizers in the early 1950s
- 1952: Bell Labs built a statistical system that could recognize any of the 10 digits from a single speaker (Davis+ 1952)

Áttekintés

1. Language, Thought, and Understanding

2. Az NLP története
   - A két tábor: szimbolikusok és szochasztikusok, 1957–1970
   - Four Paradigms: 1970–1983

1957–1970: Szimbolikusok I

The symbolic paradigm took off from two lines of research

- Chomsky and others on formal language theory and generative syntax
  - throughout the late 1950s and early to mid 1960s
  - many linguistics and computer scientists on parsing algorithms, initially top-down and bottom-up and then with dynamic programming
  - One of the earliest complete parsing systems was Zelig Harris's Transformations and Discourse Analysis Project (TDAP)
  - implemented between June 1958 and July 1959 at the University of Pennsylvania (Harris, 1962)
  - essentially implemented as a cascade of FSTs

1957–1970: Szimbolikusok II

The symbolic paradigm took off from two lines of research

- artificial intelligence
  - summer of 1956 John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester brought together a group of researchers for a two-month workshop on what they decided to call artificial intelligence
  - AI always included a minority of researchers focusing on stochastic and statistical algorithms (including probabilistic models and neural nets)
  - major focus: reasoning and logic
  - early natural language understanding systems were built
  - simple systems in single domains mainly by a combination of pattern matching and keyword search with simple heuristics for reasoning and question-answering
  - By the late 1960s, more formal logical systems were developed

1957–1970: Szochasztikusok I

- stochastic: in departments of statistics and of electrical engineering
- By the late 1950s, the Bayesian method was beginning to be applied to the problem of optical character recognition
- Bledsoe and Browning (1959) built a Bayesian text-recognition
  - a large dictionary
  - the likelihood of each observed letter sequence computed given each word in the dictionary by multiplying the likelihoods for each letter
- Mosteller and Wallace (1964) applied Bayesian methods to authorship attribution on The Federalist papers
- the first serious testable psychological models of human language processing based on transformational grammar

1957–1970: Szochasztikusok II

- the first online corpora
  - the Brown corpus of American English
    - a one-million-word collection of samples from 500 written texts from different genres (newspaper, novels, non-fiction, academic, etc.), which was assembled at Brown University in 1963–64 (Kučera and Francis, 1967; Francis, 1979; Francis and Kučera, 1982)
  - William S. Y. Wang's 1967 DOC (Dictionary on Computer), an on-line Chinese dialect dictionary

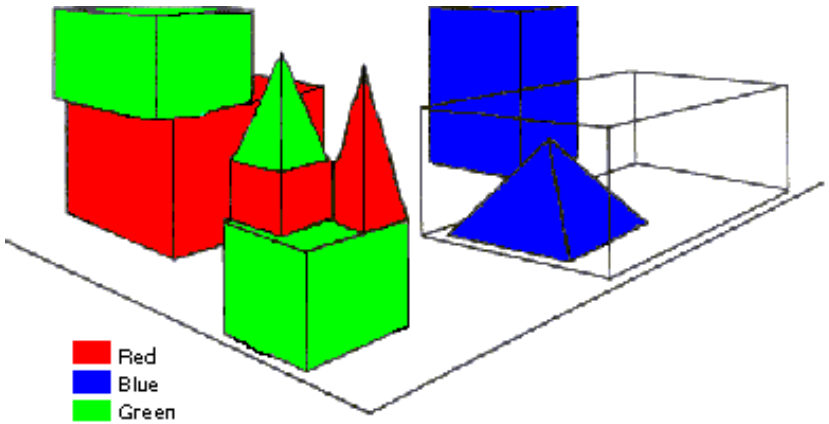
Áttekintés

Four Paradigms I, II: 1970–1983

- The stochastic paradigm
  - huge role in the development of speech recognition algorithms in this period, particularly the use of the
  - hidden Markov model (HMM) and
  - the metaphors of the noisy channel and decoding,
- The logic-based paradigm
  - begun by the work of Colmerauer+ on Q-systems and metamorphosis grammars (Colmerauer, 1970, 1975), the forerunners of Prolog, and Definite Clause Grammars (Pereira and Warren, 1980)
  - Independently, Kay's (1979) work on functional grammar
  - Bresnan and Kaplan's (1982) work on Lexical Functional Grammar (LFG), established the importance of feature structure unification

The natural language understanding field I

- Winograd's SHRDLU system
  - simulated a robot embedded in a world of toy blocks (Winograd, 1972a)
  - accepted natural-language text commands
  - Move the red block on top of the smaller green one
  - of a hitherto unseen complexity and sophistication
  - first to attempt to build an extensive (for the time) grammar of English, based on Halliday's systemic grammar
  - made it clear that the problem of parsing was well enough understood to begin to focus on semantics and discourse



The natural language understanding field II

- Roger Schank and his colleagues and students
  - the Yale School
  - a series of language-understanding programs
  - conceptual knowledge such as scripts, plans, and goals, and human memory organization (Schank and Abelson, 1977; Schank and Riesbeck, 1981; Cullingford, 1981; Wilensky, 1983; Lehnert, 1977)
  - network-based semantics (Quillian, 1968; Norman and Rumelhart, 1975; Schank, 1972; Wilks, 1975c, 1975b; Kintsch, 1974)
  - incorporate Fillmore's notion of case roles (Fillmore, 1968) into their representations (Simmons, 1973)
- The logic-based and natural-language-understanding paradigms were unified in systems that used predicate logic as a semantic representation, such as the LUNAR question-answering system (Woods, 1967, 1973)

The discourse modeling paradigm

- Grosz+: substructure in discourse, and of discourse focus (Grosz, 1977a; Sidner, 1983)
- automatic reference resolution (Hobbs, 1978)
- BDI (Belief-Desire-Intention) framework for logic-based work on speech acts (Perrault and Allen, 1980; Cohen and Perrault, 1979)

Empiricism and Finite State Models: 1983–1993

- the return of two classes of models
  that had lost popularity in the late 1950s and early 1960s,
  partially due to theoretical arguments against them such as Chomsky's influential review of Skinner's Verbal Behavior (Chomsky, 1959b)
- finite-state models
  - finite-state phonology and morphology by Kaplan and Kay (1981)
  - finite-state models of syntax by Church (1980)
- the "return of empiricism"
  - probabilistic models throughout speech and language processing,
    - IBM Thomas J. Watson Research Center on probabilistic models of speech recognition
  - probabilistic methods and other such data-driven approaches spread from speech into part-of-speech tagging, parsing and attachment ambiguities, and semantics
  - model evaluation
    - held-out data
    - quantitative metrics for evaluation
    - comparison of performance on these metrics with previous research
  - considerable work on natural language generation

The Field Comes Together: 1994–1999

- probabilistic and data-driven models become quite standard throughout NLP
  - Algorithms for parsing, part-of-speech tagging, reference resolution, and discourse processing
    all began to incorporate probabilities
  - evaluation methodologies borrowed from speech recognition and information retrieval
- increases in the speed and memory of computers $\rightarrow$
- commercial exploitation of many subareas
- especially speech recognition, and spelling and grammar correction
- algorithms began to be applied to Augmentative and Alternative Communication (AAC)
- the rise of the Web $\rightarrow$ need for language-based information retrieval and information extraction

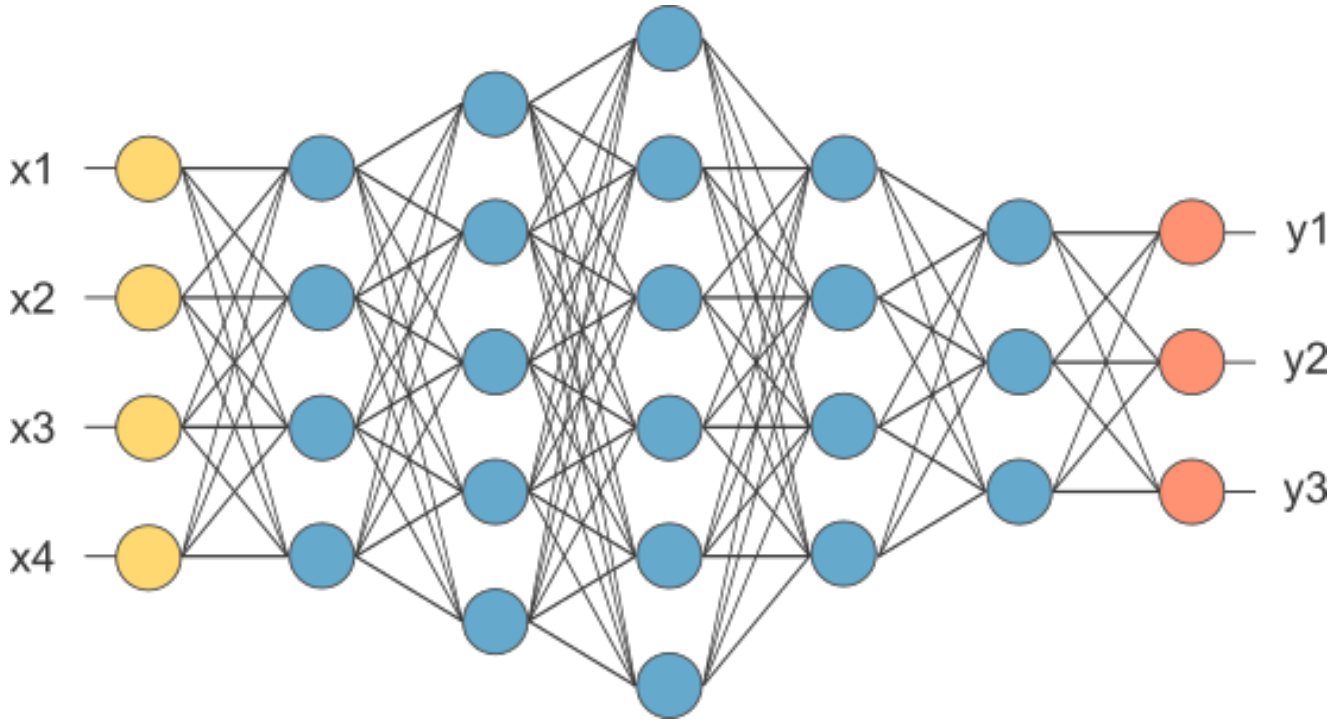
The Rise of Machine Learning: 2000“–2007” I

- empiricist trends accelerated
- largely driven by three synergistic trends
- large amounts of spoken and written material became widely available
  - the Linguistic Data Consortium (LDC) and other similar organizations
  - Penn Treebank (Marcus+ 1993), Prague Dependency Treebank (Hajič, 1998), PropBank (Palmer+ 2005), Penn Discourse Treebank (Miltsakaki+ 2004b), RSTBank (Carlson+ 2001) and TimeBank (Pustejovsky+ 2003b)
    - various forms of syntactic, semantic, and pragmatic annotations
  - casting more complex traditional problems, as problems in supervised machine learning
    - such as parsing and semantic analysis
  - additional competitive evaluations for parsing, information extraction, word sense disambiguation, question answering, summarization

The Rise of Machine Learning: 2000“–2007” II

- more serious interplay with the statistical machine learning community
  - support vector machines (Boser+ 1992; Vapnik, 1995),
  - maximum entropy techniques
    - and their equivalent formulation as multinomial logistic regression (Berger+ 1996)
  - graphical Bayesian models (Pearl, 1988)
    became standard practice in computational linguistics
- the widespread availability of high-performance computing systems
  - $\rightarrow$ the training and deployment of systems that could not have been imagined a decade earlier
- largely unsupervised statistical approaches
  - statistical approaches to
    * machine translation (Brown+ 1990; Och and Ney, 2003) and
    * topic modeling (Blei+ 2003)
  - cost and difficulty of producing reliably annotated corpora
    - became a limiting factor in the use of supervised approaches

Mesterséges ideghálók (neurális hálók)



- kibernetika (1949), konnekcionizmus (1974), mély tanulás (2006)
- feature-ök tanulása, egyre absztraktabb rétegek
-
  - gépi látás (Krizhevsky and Sutskever, 2012)
  - beszédfelismerés (Hinton et al., 2012)
- gyors tanulás a grafikus kártyán
- mint az agyban?

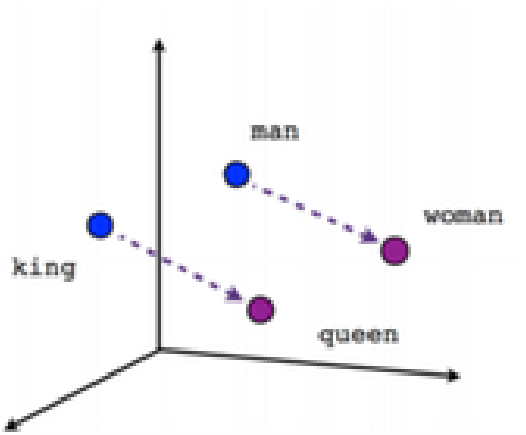
Szóbeágyazások (*word embeddings*)

- szavak reprezentációja idegh\u00e1lókban
- $\mathbf{w} \in \mathbb{R}^{300}$
- hasonló eloszlású szó $\mapsto$ közeli pont
  - eloszlás: milyen kontextusokban milyen gyakori
- milliárd szavas korpuszokon tanítjuk (*giga-word corpus*), pl. Halácsy et al., 2004; Oravecz, Váradi, and Sass, 2014
- word2vec, skip-gram, continuous bag of words (Mikolov et al., 2013)
- közös reprezentáció (Collobert et al., 2011; Hashimoto et al., 2017)
- kompozíció: karakter, morféma, szó, query, mondat, retorika
  - morfémák (Lazaridou et al., 2013)
  - szó alatti: fastText (Bojanowski et al., 2017)
  - gondolatvektor (Vaswani et al., 2017)
- kontextualizált szóbeágyazások, előtanított nyelvmodellek (Peters et al., 2018; Devlin et al., 2018; Howard and Ruder, 2018)

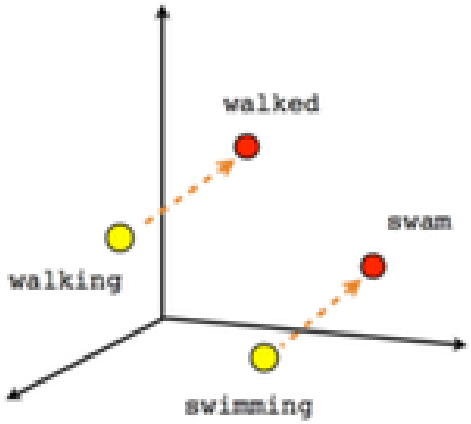
Jelentésfelbontás vektorokkal

(Mikolov, Yih, and Zweig, 2013)

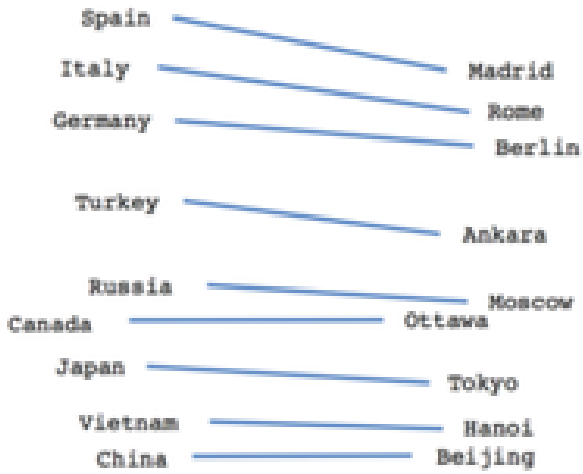
Katz and Fodor, 1963



Male-Female



Verb tense

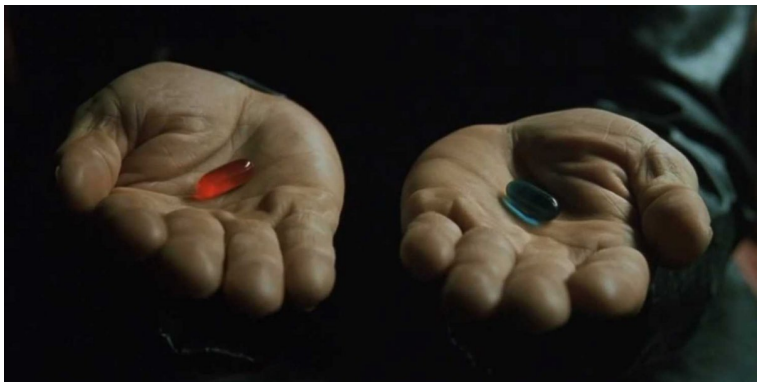


Country-Capital

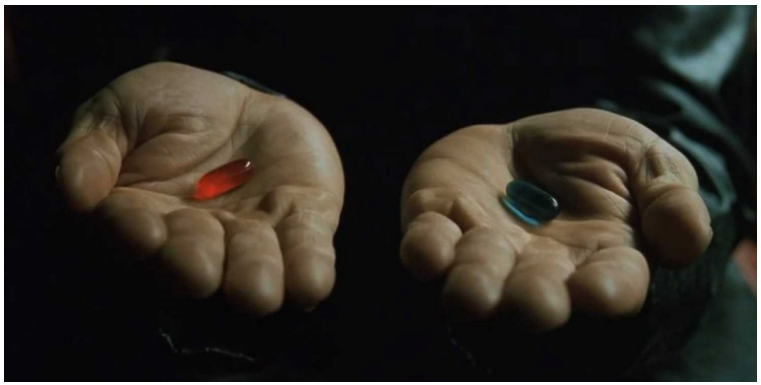
$$\textbf{király} + \textbf{nő} - \textbf{férfi} \approx \textbf{királynő}$$

- legközelebbi szomszéd (NN)

Szavak fordítása (Mikolov, Le, and Sutskever, 2013)

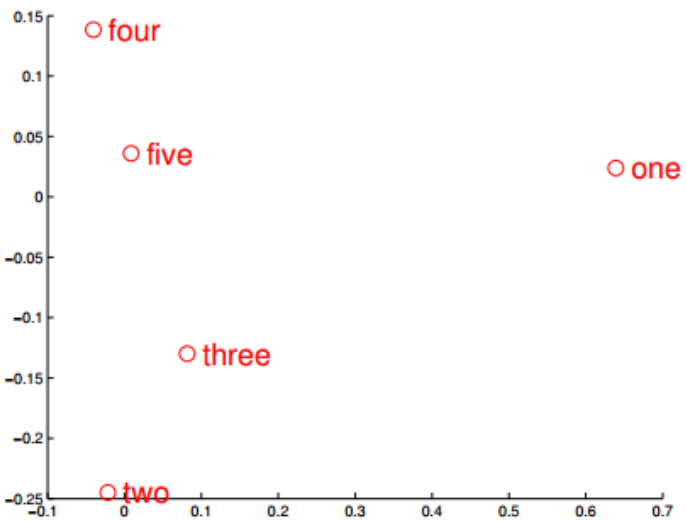


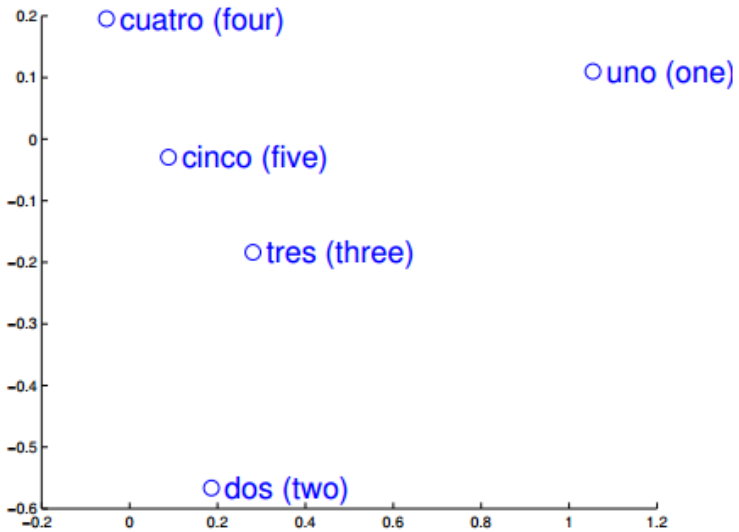
Szavak fordítása (Mikolov, Le, and Sutskever, 2013)



- két vektortér között a legegyszerűbb leképezés a lineáris (mátrix)
- most két nyelv beágyazása között
- a fordítási leképezés tanítása néhány ezer szópáron + teszt

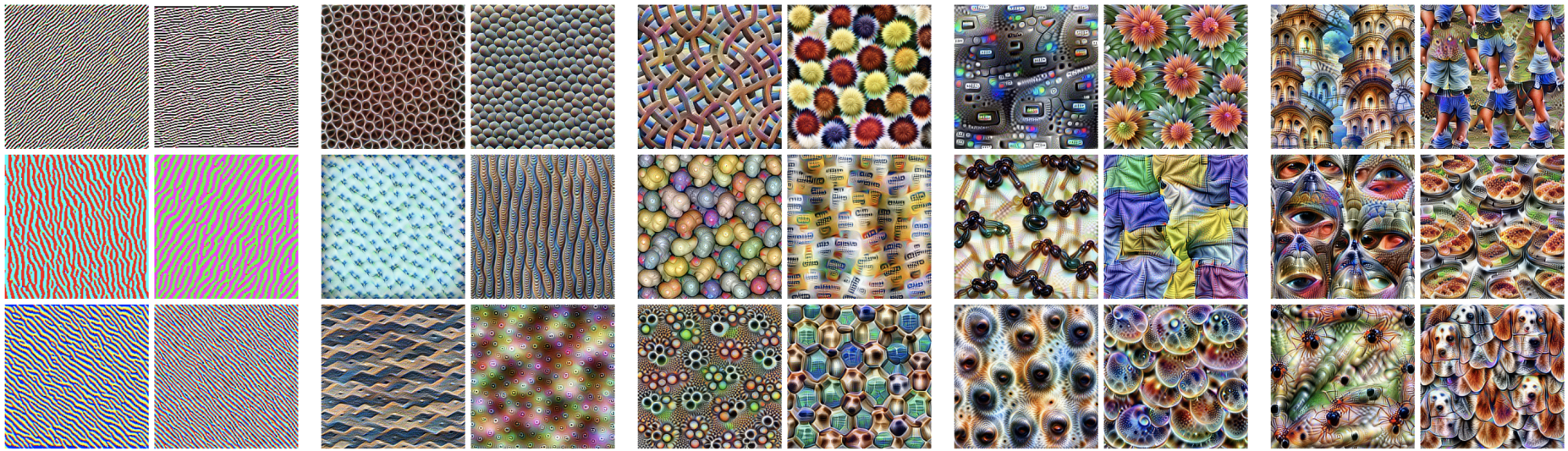
$$\min_W \sum_i ||Wx_i - z_i||^2$$





előtanított mély nyelvmodellek

aka. kontextualizált szóbeágyazások (**Rogers:2020**; Peters et al., 2018)



vision	edges $\rightarrow$ textures $\rightarrow$ patterns $\rightarrow$ parts $\rightarrow$ objects
language	morphology $\rightarrow$ syntax $\rightarrow$ semantics

- soknyelvű modellek, mondatenkóderek

Pszichológia(i realitás)

- “Many of the chapters in this book include short summaries of psychological research on human processing.”
- understanding human language processing is a scientific goal in its own right
- part of the general field of cognitive science
- it can often be helpful in building better machine models of language

Pszichológia(i realitás)

- “Many of the chapters in this book include short summaries of psychological research on human processing.”
- understanding human language processing is a scientific goal in its own right
- part of the general field of cognitive science
- it can often be helpful in building better machine models of language
- popular “wisdom”
  - direct mimicry of nature’s algorithms is rarely useful in engineering applications
  - eg if we copied nature exactly, airplanes would flap their wings; yet airplanes with fixed wings are a more successful engineering solution

Pszichológia(i realitás)

- “Many of the chapters in this book include short summaries of psychological research on human processing.”
- understanding human language processing is a scientific goal in its own right
- part of the general field of cognitive science
- it can often be helpful in building better machine models of language
- popular “wisdom”
  - direct mimicry of nature’s algorithms is rarely useful in engineering applications
  - eg if we copied nature exactly, airplanes would flap their wings; yet airplanes with fixed wings are a more successful engineering solution
- human-centered tasks $\leftrightarrow$ Airplane flight has different goals from birds
- speech recognition: exactly the task that human court reporters perform every day
- people already do this well, we can learn from nature
- for human-computer interaction
  - it makes sense to copy a solution that behaves the way people are accustomed to