



UNIVERSITÄT
ZU KÖLN

Machine Learning, Part 3, Linguistic Overview

Einführung in die Informationsverarbeitung

Nils Reiter

November 16, 2023

Recap: Neural Networks

- ▶ Many individual neurons in combination
- ▶ During training weights will be adapted, until best approximation to training data is reached
- ▶ Prediction: Application of weights, i.e., multiplication and addition
- ▶ Word2Vec
 - ▶ Use a neural network to predict whether two words are really appearing together in texts
 - ▶ Extract learned weights as embeddings
 - ▶ Leads to embeddings that are similar, if the words appear in similar contexts
 - ▶ E.g., Berlin and Paris appear in the contexts of country capital

Example: Handwritten Digit Recognition



- ▶ Publicly available data set
- ▶ Scans of hand-written digits, with correct labels
 - ▶ E.g., second row represents the digit 2
- ▶ Each image is 28x28 pixels of grey scale values

demo

Section 1

Language and Linguistics

Language and Linguistics

Phonology and Phonetics

Morphology

Syntax

Semantics

Pragmatics

What is Linguistics?

Linguistics is the scientific study of language.

Wikipedia, 925699120

What is Linguistics?

Linguistics is the scientific study of language.

Wikipedia, 925699120

- ▶ 'Scientific study'
 - ▶ 'the' scientific method
 - ▶ Testable explanations
- ▶ Language
 - ▶ ?

What is Linguistics?

Linguistics is the scientific study of language.

Wikipedia, 925699120

- ▶ 'Scientific study'
 - ▶ 'the' scientific method
 - ▶ Testable explanations
- ▶ Language
 - ▶ ?

Prescriptive vs. descriptive

- ▶ Prescriptive: Telling people how to use language
- ▶ Descriptive: Observing and analysing how people do use language

What is Linguistics?

Linguistics is the scientific study of language.

Wikipedia, 925699120

- ▶ 'Scientific study'
 - ▶ 'the' scientific method
 - ▶ Testable explanations
- ▶ Language
 - ▶ ?

Prescriptive vs. descriptive

- ▶ Prescriptive: Telling people how to use language
- ▶ Descriptive: Observing and analysing how people do use language
- ▶ Academic linguistics: Nowadays mostly descriptive

What is Language?

- ▶ Communication system
- ▶ Conventionalised: We agree (mostly)
 - ▶ Only partially authoritative

What is Language?

- ▶ Communication system
- ▶ Conventionalised: We agree (mostly)
 - ▶ Only partially authoritative
- ▶ What do we agree on?
 - ▶ Relation between *signs* and its *meaning* (which is not the same!)
 - ▶ Saussure: Semiotics
 - ▶ E.g.: 'the students in this class' *means* all of you

Linguistic sign

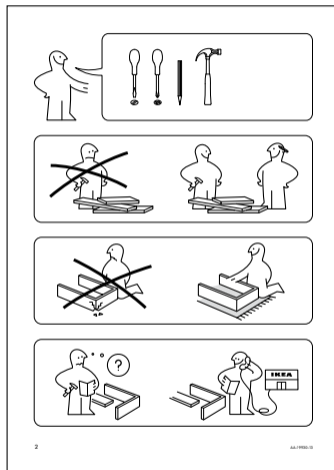
- ▶ Abstract notion
- ▶ Different levels
 - ▶ Texts
 - ▶ Sentences
 - ▶ Words
 - ▶ Syllables, morphemes
 - ▶ Spoken utterances

Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
 - ▶ Texts
 - ▶ Sentences
 - ▶ Words
 - ▶ Syllables, morphemes
 - ▶ Spoken utterances
 - ▶ Non-textual signs
 - ▶ Emojis 🥰

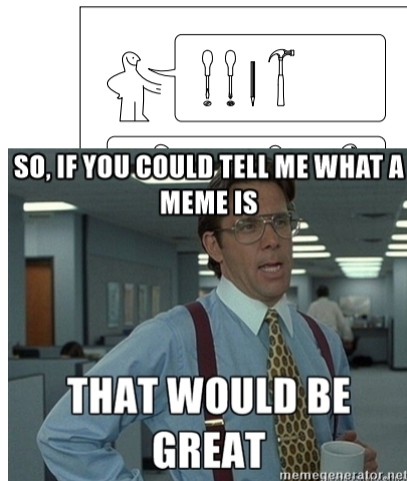
Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
 - ▶ Texts
 - ▶ Sentences
 - ▶ Words
 - ▶ Syllables, morphemes
 - ▶ Spoken utterances
 - ▶ Non-textual signs
 - ▶ Emojis 🥰
 - ▶ Assembly instructions



Linguistic sign

- ▶ Abstract notion
- ▶ Different levels
 - ▶ Texts
 - ▶ Sentences
 - ▶ Words
 - ▶ Syllables, morphemes
 - ▶ Spoken utterances
 - ▶ Non-textual signs
 - ▶ Emojis 🥰
 - ▶ Assembly instructions
 - ▶ Memes
 - ▶ ...



Linguistic sign

Meaning is arbitrary

- ▶ 'dog' refers to a four-legged, wolf-like mammal



Linguistic sign

Meaning is arbitrary

- ▶ 'dog' refers to a four-legged, wolf-like mammal
- ▶ This is an arbitrary decision
- ▶ The sign could be different, e.g., 'cat'
- ▶ No inherent meaning in signs
 - ▶ ...but strongly conventionalised



Linguistic sign

Meaning is arbitrary

- ▶ 'dog' refers to a four-legged, wolf-like mammal
- ▶ This is an arbitrary decision
- ▶ The sign could be different, e.g., 'cat'
- ▶ No inherent meaning in signs
 - ▶ ...but strongly conventionalised
- ▶ Interpreting signs (and language) is something we learn
- ▶ Language is a social construct
 - ▶ Studying language is different from studying gravity



Ambiguities

- ▶ Der Jäger traf den Mann mit dem Gewehr.
- ▶ Mädchenhandelsschule
- ▶ Maria hat Petra beim Einkaufen getroffen. Sie hat ihr Geld geliehen.
- ▶ maria hat dort liebe genossen.
- ▶ ...

Linguistics: Let's explain / represent / reproduce these ambiguities

Language and Linguistics

Phonology and Phonetics

Morphology

Syntax

Semantics

Pragmatics

Phonology and Phonetics

Phonetics

- ▶ How are language sounds produced and understood/processed?
- ▶ Focus: Practical, verbal and gestural use of language
- ▶ Links to biology, acoustics

Phonology

- ▶ Which function have certain phonemes within a language?
- ▶ Focus: Relation to other areas of linguistics and grammar
- ▶ Abstraction over concrete phonemes

Understanding Spoken Language

Relevant and irrelevant differences

- ▶ [ʃa:l] vs. [ʃal] (Schal vs. Schall)
 - ▶ Vowel length indicates a difference in meaning

Understanding Spoken Language

Relevant and irrelevant differences

- ▶ [ʃa:l] vs. [ʃal] (Schal vs. Schall)
 - ▶ Vowel length indicates a difference in meaning
- ▶ [ro:t] vs. [RO:t] (rot)
 - ▶ Pronunciation of /r/ doesn't make a difference (in German)

Understanding Spoken Language

Relevant and irrelevant differences

- ▶ [ʃa:l] vs. [ʃal] (Schal vs. Schall)
 - ▶ Vowel length indicates a difference in meaning
- ▶ [ro:t] vs. [RO:t] (rot)
 - ▶ Pronunciation of /r/ doesn't make a difference (in German)

International Phonetic Alphabet (IPA)

- ▶ <https://www.internationalphoneticassociation.org>
- ▶ Symbols defined via physiological properties of the pronunciation

Sächsische Kundin bucht Bordeaux statt Porto

Eine undeutliche Aussprache im Reisebüro kann teuer werden. Fast 300 Euro muss eine Kundin aus Sachsen für einen Flug zahlen, den sie nie angetreten hat - weil sie den gewünschten Zielort Porto dialektbedingt nicht klar artikulierte.



Corbis

Bordeaux in Frankreich: Verwechslung mit anderen Städten nicht ausgeschlossen

Pronunciation mishaps

Bordeaux vs. Porto

- ▶ Porto: ['pɔʁto]
- ▶ Bordeaux: [bɔʁ'doː]

Pronunciation mishaps

Bordeaux vs. Porto

- ▶ Porto: [ˈpɔʁto]
- ▶ Bordeaux: [bɔʁˈdoːʁ]
- ▶ Key difference: Voicing of the plosives p/b and t/d
 - ▶ /p/, /t/: voiceless (stimmlos)
 - ▶ /b/, /d/: voiced (stimmhaft)

Pronunciation mishaps

Bordeaux vs. Porto

- ▶ Porto: [ˈpɔʁto]
- ▶ Bordeaux: [bɔʁˈdoːʁ]
- ▶ Key difference: Voicing of the plosives p/b and t/d
 - ▶ /p/, /t/: voiceless (stimmlos)
 - ▶ /b/, /d/: voiced (stimmhaft)

Voice and Plosives

- ▶ Voice
 - ▶ Sounds with the use of the larynx (dt. Stimmlippen)
 - ▶ Example: Phase (voiceless: /f/) vs. Vase (voiced: /v/)
 - ▶ You can feel voice if you touch your throat

Pronunciation mishaps

Bordeaux vs. Porto

- ▶ Porto: [ˈpɔʁto]
- ▶ Bordeaux: [bɔʁˈdoːʁ]
- ▶ Key difference: Voicing of the plosives p/b and t/d
 - ▶ /p/, /t/: voiceless (stimmlos)
 - ▶ /b/, /d/: voiced (stimmhaft)

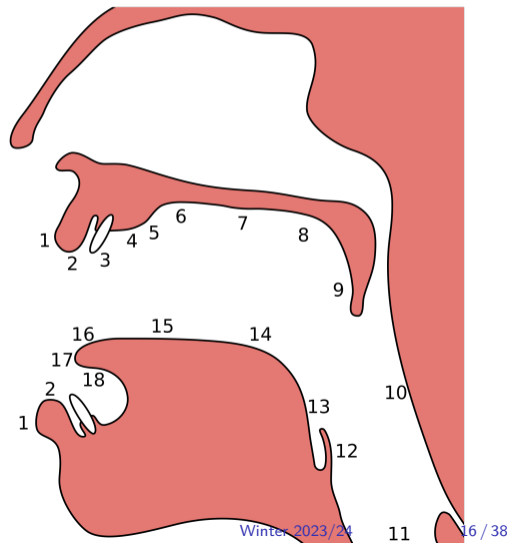
Voice and Plosives

- ▶ Voice
 - ▶ Sounds with the use of the larynx (dt. Stimmlippen)
 - ▶ Example: Phase (voiceless: /f/) vs. Vase (voiced: /v/)
 - ▶ You can feel voice if you touch your throat
- ▶ Plosive
 - ▶ Air stream is blocked, but suddenly re-opened
 - ▶ Example: /bʊs/ (plosive) vs. /mʊs/ (nasal)

Producing Sounds

Important Locations for German Sounds (Consonants)

2. labial (Lippen): [b], [p]
3. dental (Zähne): [v], [f]
4. alveolar (Zahnfach): [d], [t], ...
5. postalveolar: [ʃ]
7. palatal: [ç]
8. velar: [g], [k], ...
11. glottal: [ʔ]
 - ▶ 'ein Echo': [am ʔεçɔ]
 - ▶ 'Student:in': [ʃtu'dentʔɪn]



Producing Sounds

Consonants vs. Vowels

- ▶ Consonant
 - ▶ Produced with (complete or partial) closure of the vocal tract
 - ▶ labial/dental/... describes the position of the closure in the tract

Producing Sounds

Consonants vs. Vowels

- ▶ Consonant
 - ▶ Produced with (complete or partial) closure of the vocal tract
 - ▶ labial/dental/... describes the position of the closure in the tract
- ▶ Vowel
 - ▶ Produced without closure of the vocal tract
 - ▶ Usually voiced
 - ▶ Shaped by tongue position and lip rounding
 - ▶ (this is a simplification)

Subsection 2

Morphology

Language and Linguistics

Phonology and Phonetics

Morphology

Syntax

Semantics

Pragmatics

Morphology

- ▶ How do we create words?






Morphology

- ▶ How do we create words?
- ▶ Ambiguity:
 - ▶ Order in which parts of words are assembled

Morphology

- ▶ How do we create words?
- ▶ Ambiguity:
 - ▶ Order in which parts of words are assembled
- ▶ Morphological processes are language-dependent
 - ▶ German: Nominal composition
 - ▶ Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

References I

-  Austin, John Langshaw (1962). *How to Do Things with Words*. William James lectures. Harvard University Press.
-  Bar-Hillel, Yehoshua (1971). “Out of the pragmatic wastebasket”. In: *Linguistic Inquiry* 2, pp. 401–407.
-  Davidson, Donald (1967). “Truth and meaning”. In: *Synthese* 17.1, pp. 304–323.
-  Grice, Herbert Paul (1975). “Logic and conversation”. In: *Syntax and Semantics* 3.S 41. Ed. by P. Cole/J. Morgan, p. 58.
-  Levinson, Stephen C. (1983). *Pragmatics*. Cambridge University Press.