

Kommunikationssysteme und Spracheigenschaften

Katja Markert

Institut für Computerlinguistik
Uni Heidelberg
markert@cl.uni-heidelberg.de

October 31, 2022

- 1 Definitionen von Sprache und Kommunikation
- 2 Sind andere Kommunikationsformen (tierische und menschliche) fundamental unterschiedlich zu menschlicher Sprache?
- 3 Welche Eigenschaften sind für menschliche Sprache typisch?
- 4 Kindlicher Spracherwerb
- 5 Wie beeinflussen solche Erkenntnisse die automatische Sprachmodellierung?

Beispiele von Kommunikationssystemen?

Definition von Kommunikationssystem

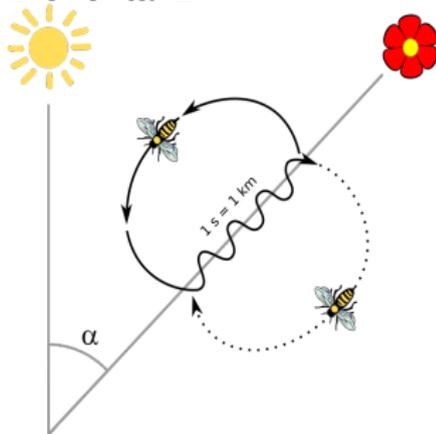
Kommunikationssysteme

”Communication consists of exchanges of information between a sender and a receiver using a code of specific signals that usually serve to meet common challenges (reproduction, feeding, protection) and, in group-living species, to promote cohesiveness of the group.”
(Vauclair, 1996, p. 99)

Kanäle:

- Visuell
- Auditorisch
- Olfaktorisch
- Taktile/Haptisch
- Biochemisch
- Elektromagnetisch
- Seismisch
- Thermal

- Bientanz



Picture from Jüppsche-commonswiki, own work, CC BY-SA 2.5

<https://creativecommons.org/licenses/by-sa/2.5>, via Wikimedia

Commons https://commons.wikimedia.org/wiki/File:Bee_dance.svg

- Meerkatzenrufe
- Mimik
- ...

Übung: Spezielle Eigenschaften menschlicher Sprache?

- **Akustisch-auditiver Kanal**
- **Übertragung und gerichteter Empfang**
- **Vergänglichkeit**

Unsere Stimmbänder haben sich dazu entwickelt, Sprache zu erlauben (Kontrast: Schimpansen)

- **Diskretheit:**

/a/

vs.



- **Arbitrarität zwischen Signifikant/Signifikat:**

/bat/ vs. ,



Picture from AnonymousUnknown author, Public domain, via Wikimedia Commons, https://commons.wikimedia.org/wiki/File:EB1911_-_Flight_-_Fig._3.jpg

- **Semantizität**

- **Dislokation:** *computer of Cleopatra, yesterday*

- **Dualität der Musterbildung:** /t/, /o/, /p/ \implies /top/,/opt/,/pot/
- **Produktivität**

Rekursion

Eine Struktur ist rekursiv, wenn sie eine Struktur derselben Art enthält.

Produktivität ist normalerweise in der Grammatik enkodiert;
Kompositionalitätsprinzip

- Austauschbarkeit in Produktion und Rezeption
- Spezialisierung für Kommunikation
- Tradierung
- Totales Feedback
- Reflexivität: “We can talk about talking”
- Lügen
- Lernbarkeit

Menschliche Sprachen erfüllen diese Kriterien, selbst wenn die Sprechersonst “kulturell” oder technologisch “weniger fortgeschritten” sind

- Menya von Papua New Guinea: ein einfaches Verb kann 2000-3000 Formen haben (Pinker, 1994)
- Cherokee Pronomensystem

- Es mag “primitive” Gesellschaften geben, aber keine primitiven Sprachen. Alle Sprachen scheinen Hocketts Charakteristika zu teilen.
- Kontrast: Es wurde noch kein tierisches Kommunikationssystem gefunden, das die wichtigsten Charakteristika menschlicher Sprache aufweist wie Rekursion, Kompositionalität, Dislokation, Grammatik. Die Suche geht aber weiter.

Andere Felder, die sich mit der Besonderheit menschlicher Sprache beschäftigen

- Kommunikation mit Tieren: Versuche Tieren menschliche Sprache beizubringen
- Sprache nicht-instinktiv, aber Fähigkeit zum Sprache lernen angeboren?: kindlicher Spracherwerb
- Neurologie: Spezielle Sprachzentren, Sprachpathologien
- Evolutionsbiologie: Sprache entstand nach Separation von anderen Primaten

Lexikon: Wortliste

- **Wort:** Lautkette, die arbiträr mit einer Bedeutung/Konzept verbunden ist
- Groß, aber endlich
- Schätzung: Wir kennen 60,000 Wörter min!
- Die meisten Forscher halten das Lexikon für nicht angeboren.
- **Wie kommen wir zu dieser Schlussfolgerung und was für Vorteile hat dies?**

Grammatik

- Kombinatorisches System
- Unendliche Anzahl von Sätzen aus endlicher, relativ kleiner Regelmenge
- Keine willkürliche Satz-Bedeutung-Verbindung
- Angeborene Elemente?



Sie lernen eine neue Sprache. Ich zeige auf dieses Bild und sage *gavagai*. Was bedeutet Gavagai? Why? Was ist wenn ich Ihnen dann noch ein weiteres Wort beibringe und wieder auf das Bild zeige und *flimt* sage?

Zwei angeborene Annahmen beim Wortlernen: Biased learning

- Whole object assumption
- Mutual exclusivity assumption

Picture from Steffen Flor, CC BY-SA 2.0

<https://creativecommons.org/licenses/by-sa/2.0>, via Wikimedia Commons, downloaded from [https://commons.wikimedia.org/wiki/File:Rabbits_of_Okunoshima,_August_2018_\(03\).jpg](https://commons.wikimedia.org/wiki/File:Rabbits_of_Okunoshima,_August_2018_(03).jpg)

Wenn ich so Worte lernen kann, dann kann ich doch so auch Sätze (=Wortketten) lernen?

Falsch!

- **Generative Grammatik:** eine Menge an Regeln, die genau spezifiziert, was ein Satz in einer Sprache sein kann oder nicht
 - *Er sitzt mir Bücher ungrammatisch Er schickt mir Bücher grammatisch*
- Bedeutungsunabhängig: *Colourless green ideas sleep furiously*
- Wahrscheinlichkeitsunabhängig: *Glowing sow gives birth to fluorescent piglets*, Spiegel, 9/1/2008
- Wahrheitsunabhängig
- Unendliche Anzahl grammatischer Sätze

Kinder acquirieren Grammatik schnell, verlässlich und ohne dass man sie explizit unterrichtet

- Poverty of the stimulus
- Implizites Regelbewusstsein \implies *This is a wug. There are two ?*
- Fehler oft regelkonform: *I holded the baby rabbits*

Gibt es eine universelle Grammatik, die einen Rahmen dafür vorgibt, wie Sprachen aussehen können? Offene Forschungsfragen!

Es gibt Evidenz, dass zumindest ein Teil der Sprachfähigkeit angeboren ist bzw. dass menschliche Sprache einzigartig ist.

- Universalität von generellen menschlichen Sprachcharakteristika
- Unterschiede zu tierischer Kommunikation
- Kindlicher Spracherwerb im Bereich Grammatik
- Physische Sprachzentren im Gehirn (siehe Appendix)
- Schwierigkeiten, Tieren menschliche Sprache beizubringen (siehe Appendix)
- Phylogenetische Entwicklung

Nativist/Rationalist

Wir müssen die Sprachregeln modellieren.

Empiristen

Wir sollten Sprache aus existierenden Texten (=Korpora) lernen.

- Reaktion auf frühe Korpuslinguistik (vor 1950), die oft nur Daten in dem gesammelten Korpus modellieren wollte
- Wollen das Wissen der Muttersprachler (=competence) modellieren. Interesse an generativer Grammatik
- Generative Grammatik: eine Menge an Regeln, die genau festsetzt, was eine grammatische Sequenz ist
- Korpus und Statistik kann nie nützlich sein. Besser Introspektion.
 - Korpus beschreibt **Performanz**: Sprachgebrauch
 - kann Kompetenz nicht modellieren
 - Korpus endlich
 - Wahrscheinlichkeit oder Vorkommen eines Satzes kann Grammatikalität nicht beschreiben

Fragen

- Ist “Ich sehe einen Mann mit einem Fernglas” a **ein grammatisch korrekter** Satz?
- *Ich sehe einen Mann mit einem fernglas. vs. Ich sehe einen Mann mit einem Bart*
- **Parsing:** Enumeriere alle mögliche Parses

Regelbasiert, symbolisch (bis Mitte 90er)

- **regelbasiert**: Regeln endlich, Sprache nicht
- Fokus auf Abdeckung **coverage**: liste alle Parses eines Satzes
- wissensintensiv
- keine Statistik
- oft manuelle Regelkreierung

- 1 *I read less books than you*
- 2 *The cat likes tuna fish*
- 3 *The cat the dog chased likes tuna fish*
- 4 *The cat the dog the rat bit chased likes tuna fish*
- 5 *The cat the dog the rat the elephant admired bit chased likes tuna fish*

- Menschen unterscheiden sich in **Grammatikalitätsurteilen**
 - *John I believe Sally said Bill believed Sue saw*
 - *The boys read Mary's story about each other*
 - *I read less books than you*
- **Akzeptabilität** statt Grammatikalität?
- Performanzbeschreibung
- NLP sollte **robust** sein
- Introspektion nicht beobachtbar
- manuelle Regelmodellierung zu aufwendig \implies knowledge acquisition bottleneck

- Disambiguierung und **wahrscheinliche Interpretationen**: *Time flies like an arrow*
- Wortfolgen “drink white ...” ?
- Probabilistische Konzepte wichtig
- Komplexe empirische Modelle gehen weit über einen Korpuslookup hinaus
- Pattern matching approaches, in particular neural networks

- Datensammlung: Korpora
- Datenrepräsentation
- Probabilistische Modelle, maschinelles Lernen oder neuronale Netzwerke
- Evaluation

Empirische Modellierung wird oft auch die linguistische Intuition des Modellierers benutzen \implies Mischformen zwischen rationalistisch und empirisch

- Fokus auf empirische oder kombinierte Modelle
- Sehr große Korpora
- Computer Power und Skalierbarkeit
- Maschinelles Lernen
- Neuronale Netze/Deep Learning

Sie sollten fähig sein:

- Hocketts Sprachcharakteristika anzuwenden
- Argumente für und gegen Sprachangeborenheit zusammenzufassen
- Behauptungen als rationalistisch oder empirisch zu kategorisieren

- Übung 1
- Hockett's Charakteristika sehr gut erklärt unter https://en.wikipedia.org/wiki/Hockett%27s_design_features

- Steven Pinker, *The language instinct* (1994), Chapter 2
- Hockett, C.: *The origin of speech*; in *Scientific American*, 203, 88-111, 1960
- Searle, John (1972): A Special Supplement: Chomsky's Revolution in Linguistics
<http://www.nybooks.com/articles/10142>
- Vauclair, J. (1996). *Animal cognition*. Cambridge, MA: Harvard University Press.
- Price et al (2015): *Vervets revisited: A quantitative analysis of alarm call structure and context specificity*. *Nature*.
<https://www.nature.com/articles/srep13220>

- Juliane Kaminski et al. (2004): *Word Learning in a Domestic Dog: Evidence for fast mapping*; in *Science* 304, 1682 (2004), available online
- Paul Bloom's answer to Kaminski: *Can a dog learn a word* <http://science.sciencemag.org/content/304/5677/1605.full>
- Griebel and Oller (2012): *Vocabulary Learning in a Yorkshire Terrier: Slow Mapping of Spoken Words*. In *PlosOne* 7(2), 2012. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0030182>
- Tempelmann, S., Kaminski, J., and Tomasello, M. (2014). *Do domestic dogs learn words based on humans' referential behaviour?*. *PloS one*, 9(3).
- John W. Pilley, Alliston K. Reid (2011): *Border collie comprehends object names as verbal referents*. In: *Behavioural Processes* 2011, available online at Science Direct, together with videos of experiments. <https://www.sciencedirect.com/science/article/abs/pii/S0376635710002925>
- Fugazza et al (2021): *Rapid learning of object names in dogs*.