

Sprache und Daten

Vortrag von Dennis Behrendt und Hannes Breuer

Im Seminar „Kreativität und digitaler Wandel“

SoSe 2017

2.5.2017

Worum geht es?

- Sprache ist elementarerer Gegenstand des Moduls
- sie spielt in verschiedenen Sinnzusammenhängen eine Rolle
 - automatische Übersetzungen
 - Spracherkennungssoftware allgemein
 - als Medium des Ideen- und Gedankenaustauschs: fast jeder kommunikative Akt wird über Sprache vermittelt
- jeder hat sein eigenes Verständnis von Sprache
- wir wollen euch heute neue Aspekte aufzeigen, die zum Arbeiten mit Sprache wichtig sind

Übersicht

- Was sind Daten allgemein?
- Was ist Sprache?
 - Ansatz der Sprachwissenschaften
 - Ansatz der Kognitionswissenschaften
- Vielfalt der Sprachen
- Was sind Sprachdaten?

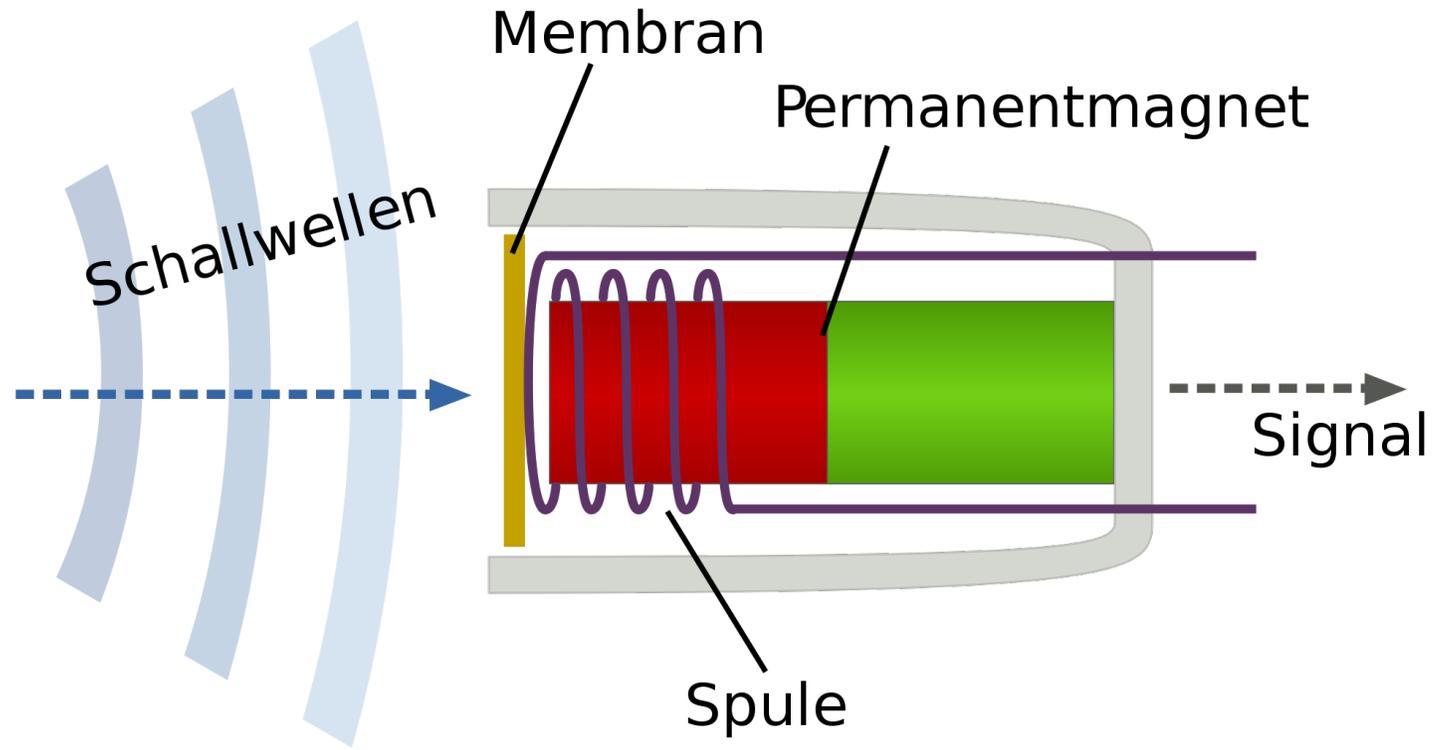
Was sind Daten?

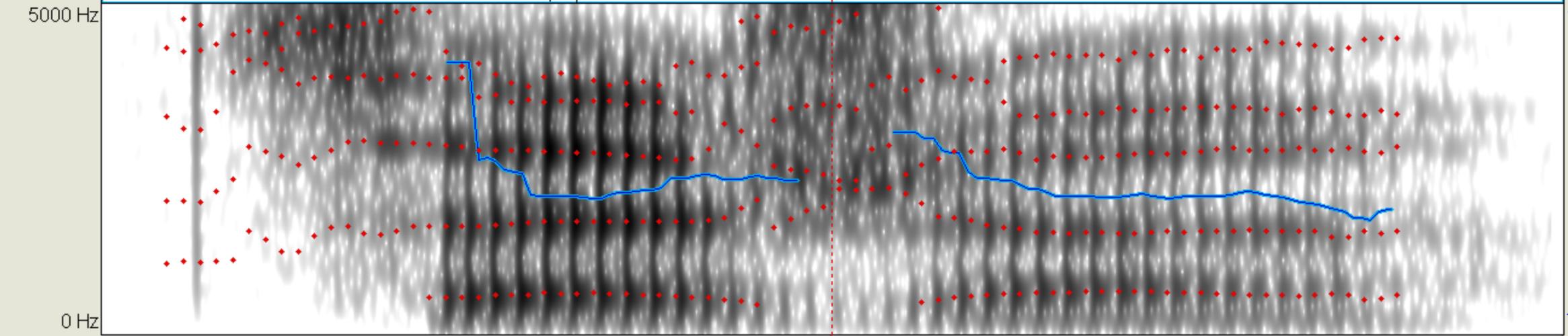
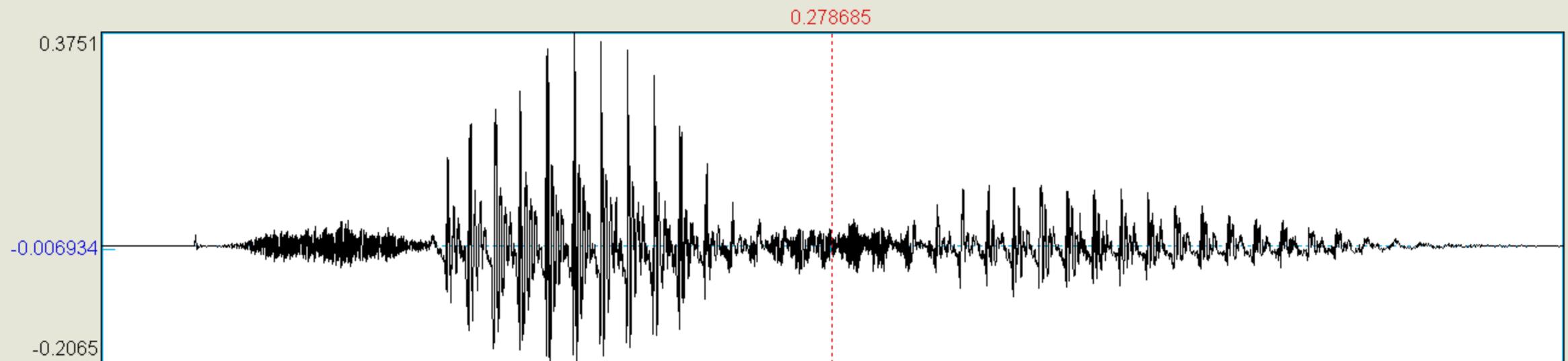
Daten als soziale Praxis

- Ziel: Soziale Praxis
 - Erhebung
 - Verwendung & Folgen der Verwendung

Digitalisierung und Daten

- Digitalisierung (Springer Wirtschaftslexikon)
 - Umwandlung, Darstellung, Durchführung von Information und Kommunikation
 - Konsequenzen für alle Lebensbereiche





0 Visible part 0.557370 seconds 0.557370

Total duration 0.557370 seconds

Digitalisierung und Daten

- Digitalisierung (Springer Wirtschaftslexikon)
 - Umwandlung, Darstellung, Durchführung von Information und Kommunikation
 - Konsequenzen für alle Lebensbereiche
- Daten
 - Grundlage für Digitalisierung
 - das zu Transformierende
 - **Daten** =
 - Angaben, (Zahlen-)Werte oder formulierbare Befunde
 - durch Messung, Beobachtung gewonnen

Wie Datenerhebung?



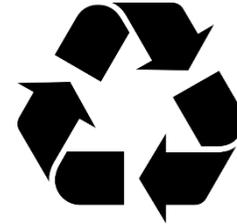
Welt



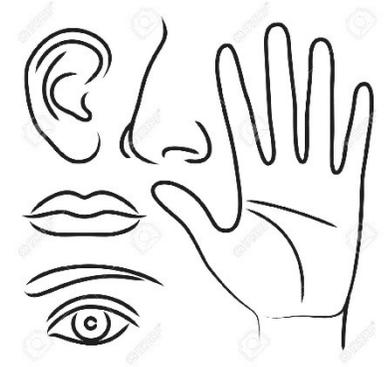
Selektion



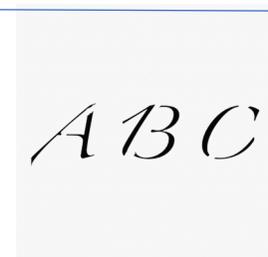
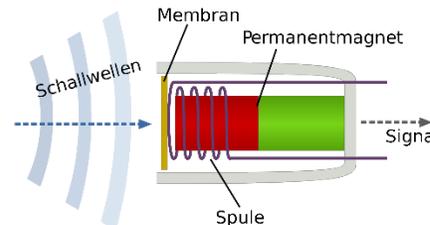
Messung



Transformation



Wahrnehmung



Ergebnis menschlichen HANDELNS



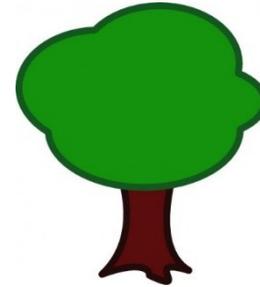
sinnhaft ↔ routiniert

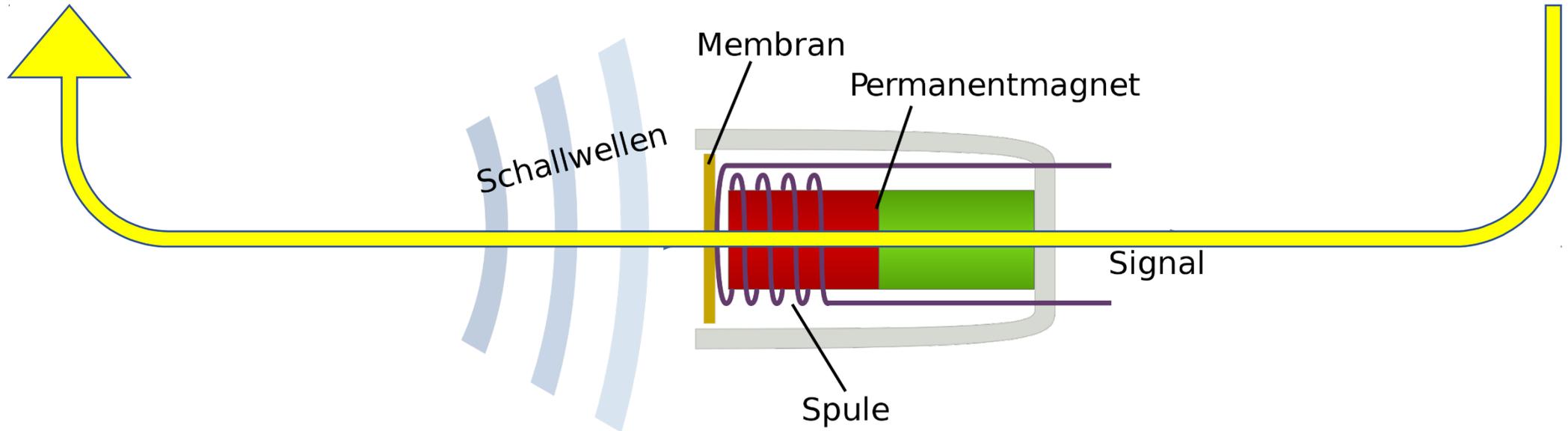
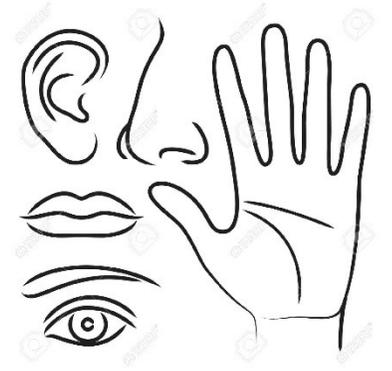
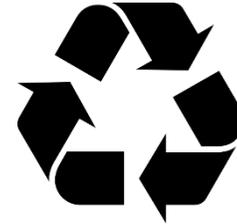


bewusst ↔ unbewusst



materiell ↔ immateriell

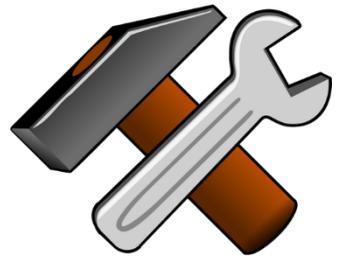




Konservieren



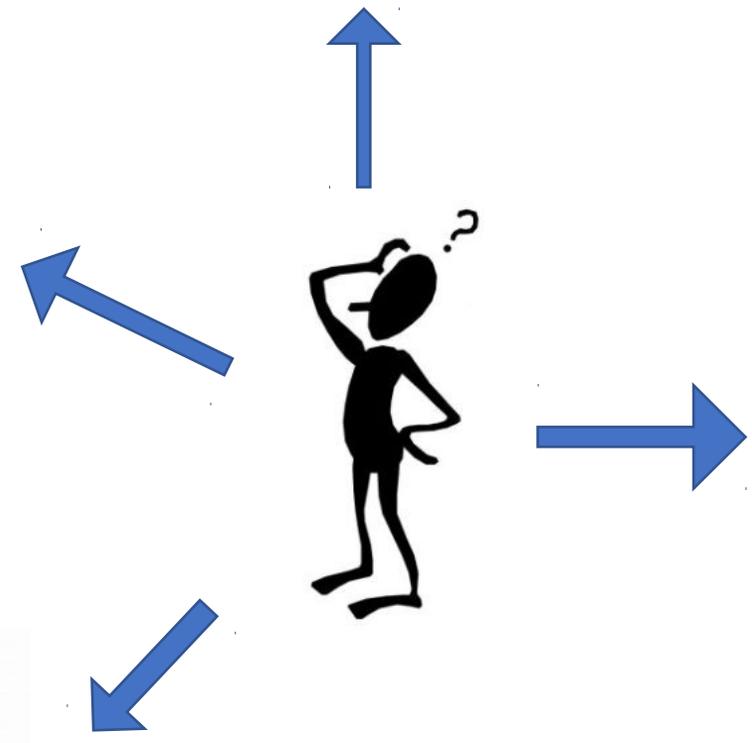
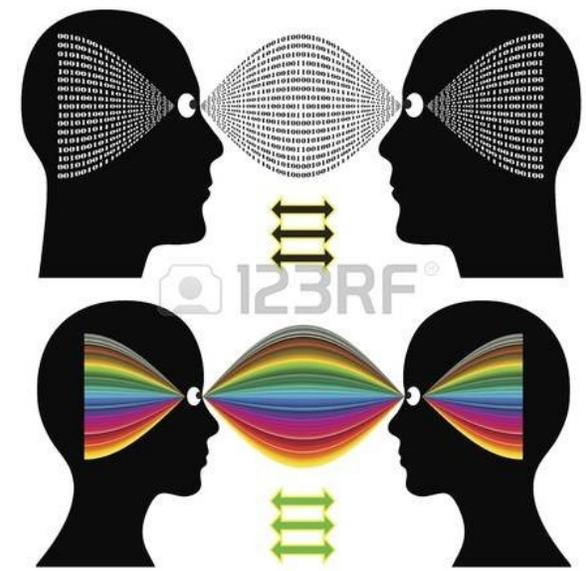
Werkzeug



Fragen



Kommunikation



Interpretation

- etwas gegebenes verstehen -

A B C

Verstehen

- Zusammenhang erkennen -

12
AIBC
14

Interpretation

=

kontextabhängig

Zwischenstand

- Bedeutung von Daten, Texten
 - selbstreferenziell
 - menschliches Konzept
 - instabil  Veränderung
- **ABER** FUNKTIONIERT trotzdem ausreichend gut für gewisse Zwecke

Was ist Sprache?

- Ansatz der Sprachwissenschaften
- Ansatz der Kognitionswissenschaften

Was ist Sprache?

Sprachwissenschaften

- Die moderne Sprachwissenschaft beruht hauptsächlich auf den Ideen von Noam Chomsky: „Universal Grammar“.
- Kernaussage: Sprache wird nicht rein „erlernt“, sondern ist bereits in Form angeborenen Sprachwissens vorhanden.
- Alle Sprachen der Welt folgen denselben Prinzipien (sog. Sprachuniversalien) und unterscheiden sich im Endeffekt nur im Wortschatz.
- Alle Sprachen der Welt können jeden Sinnzusammenhang ausdrücken.
- Sprache wird dabei als gegebenes, beobachtbares Phänomen betrachtet. Der Mensch spricht, „weil er so viel zu sagen hat“.

Was ist Sprache?

Kognitionswissenschaften

- viele Wissenschaften haben sich mit Sprache beschäftigt; hier sollen die Kognitionswissenschaften beispielhaft angeführt werden.
- Problem 1: Sprache ist nicht angeboren. Es konnten keine Evidenzen für irgendein angeborenes Sprachmodul im Gehirn gefunden werden. Bsp. ‚Sprachgen‘ FoxP2
- Problem 2: Chomskys Analysen funktionieren nur im Englischen. Für andere Sprachen müssen die Analysen komplett neu formuliert werden.
- Problem 3: Chomskys Grundannahmen sind mittlerweile teilweise widerlegt. Bsp. ‚poverty-of-input-argument‘
- Problem 4: Sprache ist kein reines Kommunikationsmittel/-werkzeug:
2 Beispiele

- **Beispiel 1: Post-traumatische Belastungsstörung**
 - Menschen mit einer PTBS haben verschiedene Gendefekte, die durch die Erkrankung ausgelöst werden.
 - Nach einer Psychotherapie, also einer Gesprächstherapie, sind die Gendefekte nicht mehr nachweisbar (Yehuda et. al. 2013).
- **Beispiel 2: Priming-Effekte**
 - Priming ist die Grundannahme, dass eine (auch sprachliche) Beeinflussung eines Menschen sein Verhalten ändert und dass diese Verhaltensänderung voraussagbar ist.
 - Experimenteller Beweis wurde von Bargh et. al. 1996 erbracht („Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action“)

Was also ist Sprache? Pathologie

- Menschen mit Sprach- und Sprechstörungen können komplexe Sachverhalte genauso verstehen und erschließen wie sprechende Menschen.
 - Bsp.: taubstumm geborene Menschen
- Bei manchen Störungen können logische Gedankengänge aber gestört sein.
 - Bsp.: Wernicke-Aphasie (Sprachverständnis gestört)

Was also ist Sprache?

- Zwar ist Sprache beobacht- und beschreibbar und sie ermöglicht es uns, ziellose Gespräche zu führen („Smalltalk“), aber sie ist gleichzeitig auch Ausdruck unserer Persönlichkeit und etwas, das uns als Mensch definiert.

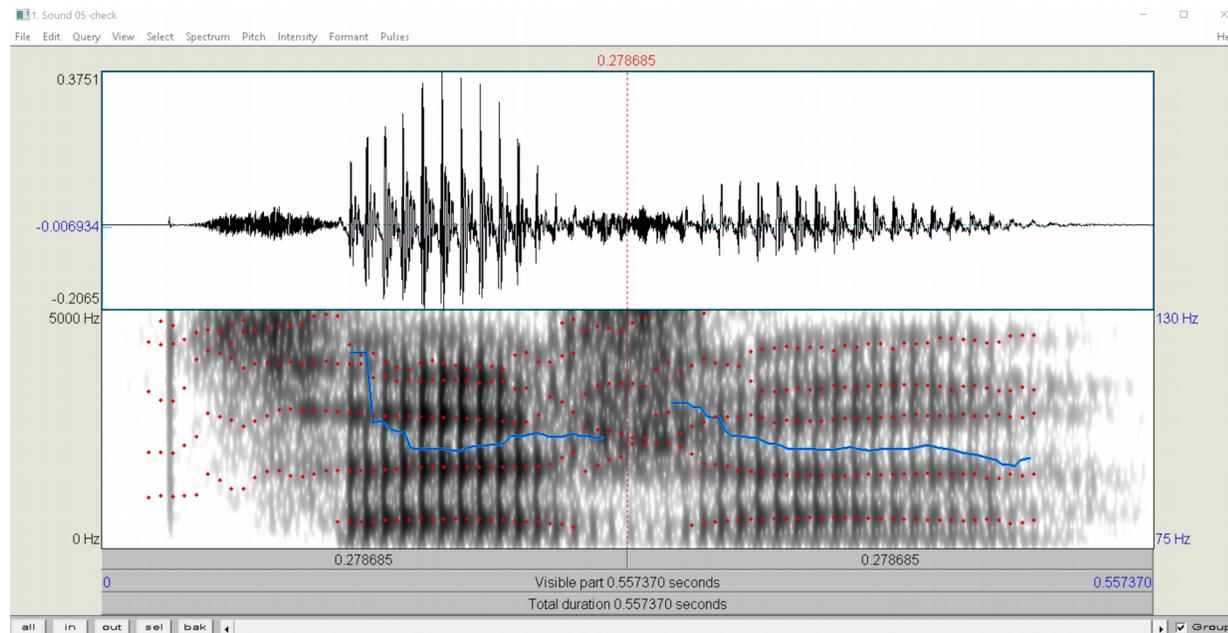
Wie kann Technik das verarbeiten?

Probleme bei der Sprachverarbeitung

- Bei der computergestützten Analyse und Verarbeitung von Sprache muss die Interpretationsfähigkeit des menschlichen Gehirns imitiert werden.
- Unsere These: ohne das Gesprochene/Geschriebene zu „verstehen“, kann kein Sinngehalt erfasst und verarbeitet werden.
- 3 Beispiele

Beispiel 1: Verarbeitung gesprochener Sprache

- ist die häufigste Anwendung der computergestützten Sprachverarbeitung im Alltag
- inzwischen sehr viele Produkte für diesen Bereich (Siri, Google Now, Cortana, ...)



Aber wie gut funktioniert das Ganze?

Bsp. 2: Ist links und rechts links und rechts?

„It is generally acknowledged that perception of and orientation in space are determinant factors in human action and interaction.

As such, speech heavily depends on knowledge of the context: Where and When is a sentence uttered, and by Whom.

These three dimensions are traditionally seen as the so-called deictic centre of all linguistic events, without which no linguistic expression can be properly interpreted.” [\(Quelle\)](#)

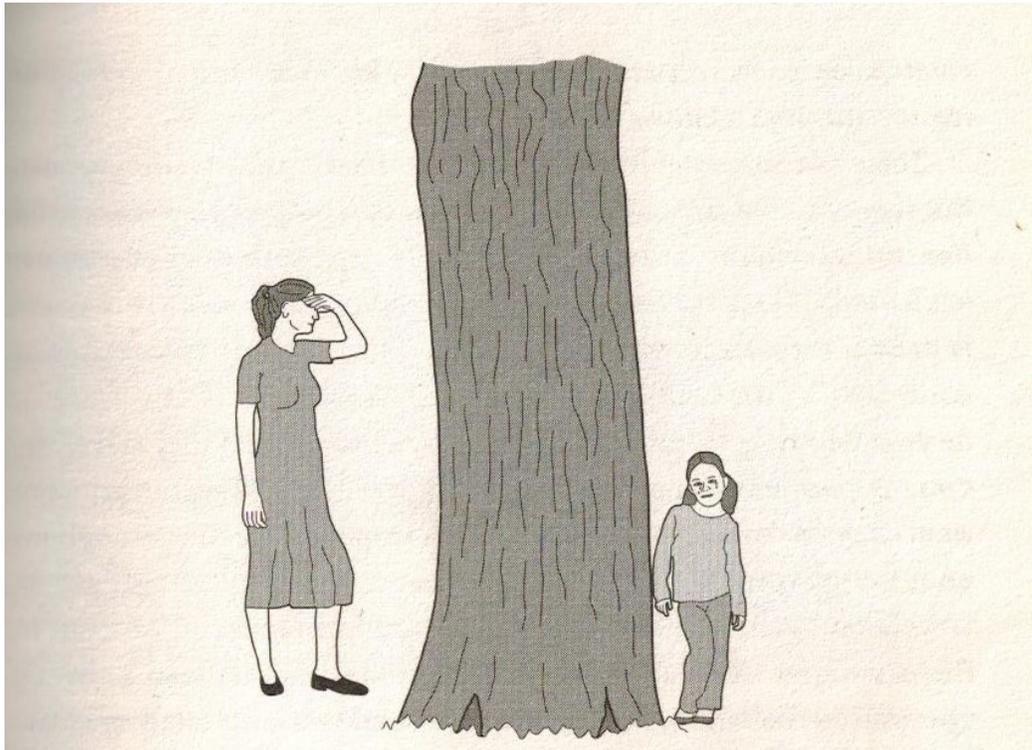
- Es scheint offensichtlich, links und rechts einfach in alle Sprachen der Welt übersetzen zu können. Aber tatsächlich ist das nicht so.

- Guugu Yimithirr ist eine Aboriginesprache Australiens.
- verwendet kein relationales Bezugssystem (links, rechts, vorne, hinten), sondern ein absolutes Bezugssystem mit Himmelsrichtungen:



Quelle:
https://theblogbyjavier.files.wordpress.com/2013/07/img_2500.jpg

Eine Bildbeschreibung auf Guugu Yimithirr



bula gabiir gabiir

zwei Mädchen

Nyulun nubuun yindu buthiil naga

eines hat die Nase nach Osten

Nyulu yindu buthiil jibaarr

die andere hat die Nase nach Süden

Yugu gaarbaarr yuulili

ein Baum steht zwischen ihnen

Buthil jibaar nyulu baajiljil

sie weint mit der Nase nach Süden

☑ Der Leser hatte das Buch nördlich von sich

- Levinson (1997):

„in order to speak Guugu Yimithirr, it is necessary to carry out a specialized kind of background computation of orientation and direction.

Further, these computations must be carried out well in advance of (indeed independently of) speaking, and the results must be memorized, be available for inference and other psychological processes, and in general pervade many aspects of cognition.”

- Deutscher (2011: 172):

“you need to know exactly where the north, south, west, and east are, since otherwise you would not be able to impart the most basic information.

It follows, therefore, that in order to be able to speak such a language, you need to have a compass in your mind, one that operates all the time, day and night, without lunchbreaks or evenings.”

Links und rechts einfach übersetzen?

- aus einer Bedienungsanleitung für einen Gasofen:

Die Kochstellen können im Bereich **<kleine Flamme>** und **<große Flamme>** stufenlos geregelt werden. Drehen im Uhrzeigersinn (=nach rechts) erhöht die Brennerleistung. Drehen gegen den Uhrzeigersinn (=nach links) verringert die Brennerleistung.

✉ Lebensgefahr!

- „In Deutschland fährt man auf der rechten Straßenseite.“
- Anzeige auf dem Handys: „Zum Entsperren nach oben wischen.“

Bsp. 3: Verwandtschaftsgrade

- Sprachen kodieren Verwandtschaftsbeziehungen zwischen Menschen auf unterschiedliche Weise.
- Nerlove & Romney (1967): Untersuchung zu Geschwisterbezeichnungen

	♂Referent		♀Referent		Beispiel
	♂Sprecher	♀Sprecher	♂Sprecher	♀Sprecher	
älter	1	3	5	7	maximal
jünger	2	4	6	8	
älter jünger	<i>brother</i>		<i>sister</i>		Englisch
älter jünger	<i>kakak</i>				Indonesisch
	<i>adik</i>				
älter jünger	<i>ani</i>		<i>ane</i>		Japanisch
	<i>otōto</i>		<i>imōto</i>		
älter jünger	<i>thabuju</i>	<i>kularrind</i>	<i>kularrind</i>	<i>yakukathu</i>	Kayardild
	<i>duujing</i>	<i>kularrind</i>	<i>kularrind</i>	<i>duujing</i>	

Es sind 4140 Bezeichnungssysteme logisch möglich, von denen in der Untersuchung aber nur 14 beobachtet werden konnten (Stichprobe von 245 Sprachen).

Bei diesen 4140 Möglichkeiten sind ausgenommen: Eltern, Großeltern, Tanten und Onkel, Cousins und Cousinen und z.B. ob angeheiratet oder nicht, uvm...

Was sind Sprachdaten?

- Allgemein: Jede Form niedergeschriebener oder gesprochener Sprache. ✉ nicht ausreichend für z.B. sinnerhaltende Übersetzungen.
- Spezieller: Jede Form niedergeschriebener oder gesprochener Sprache und alle Informationen, die zum Interpretieren gesprochener oder geschriebener Sprache notwendig sind.
- Probleme:
 - Herkunft und Dialekt des Sprechers?
 - Himmelsrichtung beim Sprech- und Schreibakt sind Sprachdaten?
 - Alter und Familienkonstellation des Sprechers sind Sprachdaten?

Das tatsächlich Gemeinte zählt zu den Sprachdaten um den Inhalt erfassen zu können?

Der Computer muss also dem Sprecher „in den Kopf schauen“ können?

Quellen

• Video

- Cambridge Analytica – The Power of Big Data and Psychographics
 - <https://www.youtube.com/watch?v=n8Dd5aVXLCc>
- TED Talk - Beau Lotto
 - https://www.ted.com/talks/beau_lotto_optical_illusions_show_how_we_see?language=de
- <https://www.youtube.com/watch?v=BbFew1T6Wck&feature=youtu.be>

• Journals/Zeitschriften

- Focus – Womit die Deutschen Ihre Lebenszeit verbringen
 - http://www.focus.de/gesundheit/videos/schlafen-kochen-putzen-womit-die-deutschen-ihre-lebenszeit-verbringen_id_4140015.html
- Forbes – Data is not the same as truth
 - <https://www.forbes.com/sites/kalevleetaru/2016/01/04/data-is-not-the-same-as-truth-interpretation-in-the-big-data-era/#3c975a442f8a>
- Yehuda et. al. „Epigenetic Biomarkers as Predictors and Correlates of Symptom Improvement following Psychotherapy in Combat Veterans with PTSD“, doi: 10.3389/fpsyt.2013.00118
- Bargh et. al. 1996 „Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action“, Journal of Personality and Social Psychology 1996. Vol. 71, No. 2. 230-244

• Buch

- Daniel Kahneman – Thinking, Fast and Slow
- Niklas Luhmann – Einführung in die Systemtheorie
- Alfred Schütz – Der sinnhafte Aufbau der sozialen Welt
- Guy Deutscher – Through the Language Glass

Danke sehr für Eure
Aufmerksamkeit!