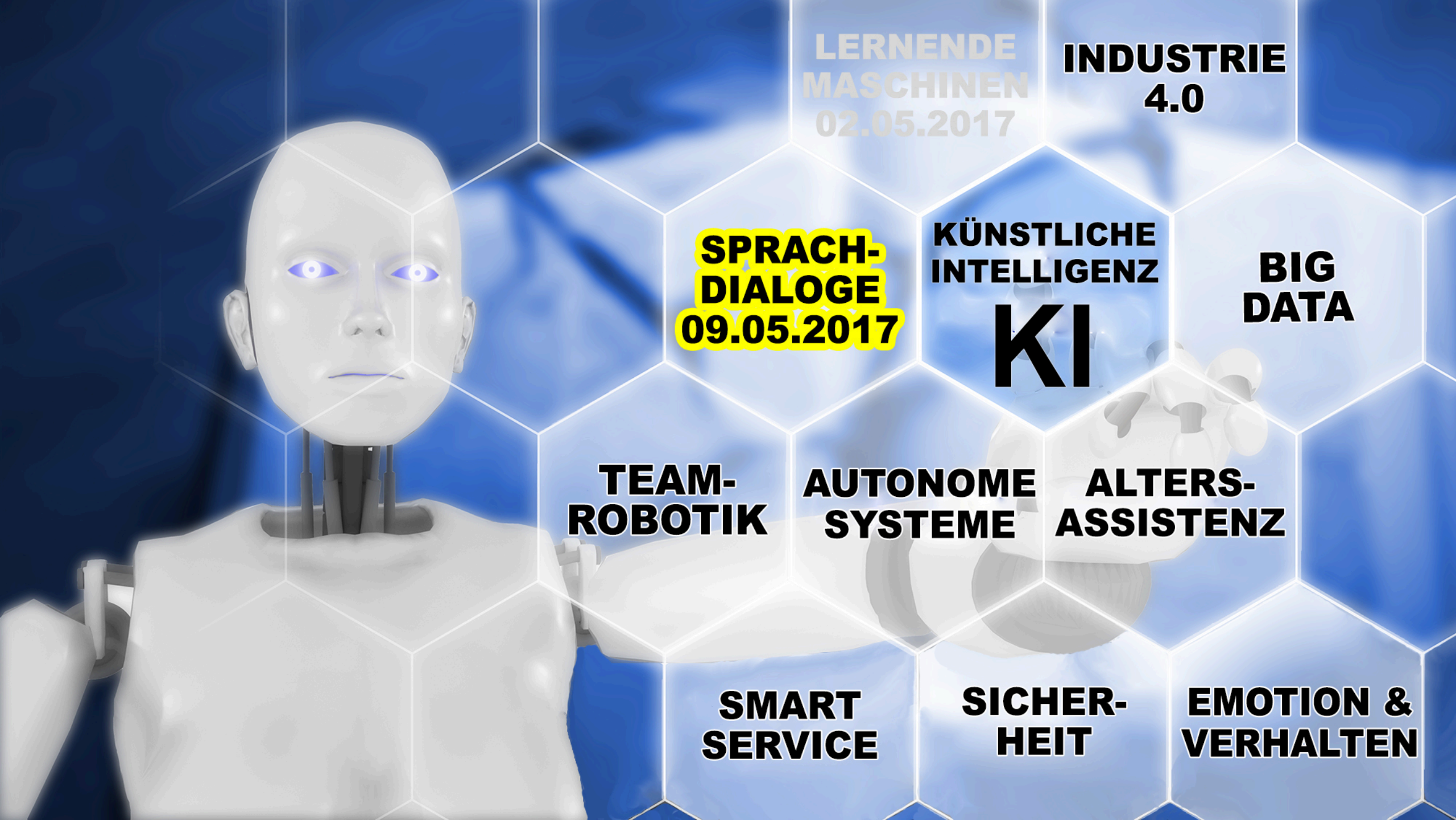

LERNENDE MASCHINEN 02.05.2017
INDUSTRIE 4.0
SPRACH-DIALOGE 09.05.2017
KÜNSTLICHE INTELLIGENZ
KI
BIG DATA
TEAM-ROBOTIK
AUTONOME SYSTEME
ALTERS-ASSISTENZ
SMART SERVICE
SICHER-HEIT
EMOTION & VERHALTEN

Online-Kurs von DFKI und acatech zum maschinellen Lernen

Im April 2017 waren bereits 4700 Teilnehmer eingeschrieben. In 3 Kurswochen bieten Wissenschaftler, Vertreter aus Unternehmen, Entwickler und Anwender in insgesamt 38 Videos Orientierungswissen für das maschinelle Lernen.

mooc.HOUSE https://mooc.house/courses/machinelearning-2016

Vorlesungsreihe 2017: Künstliche Intelligenz für den Menschen: Digitalisierung mit Verstand

Mainz, 09. Mai 2017

Johannes Gutenberg Stiftungsprofessur

Wie können Computer unsere menschliche Sprache verstehen?

Vom Sprachdialogsystem bis zum Simultandolmetscher

Prof. Dr. rer. nat. Dr. h.c. mult.

Wolfgang Wahlster

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Saarbrücken/Kaiserslautern/Bremen/Berlin/Osnabrück
Tel.: (0681) 85775-5252
E-mail: wahlster@dfki.de
www.dfki.de/~wahlster

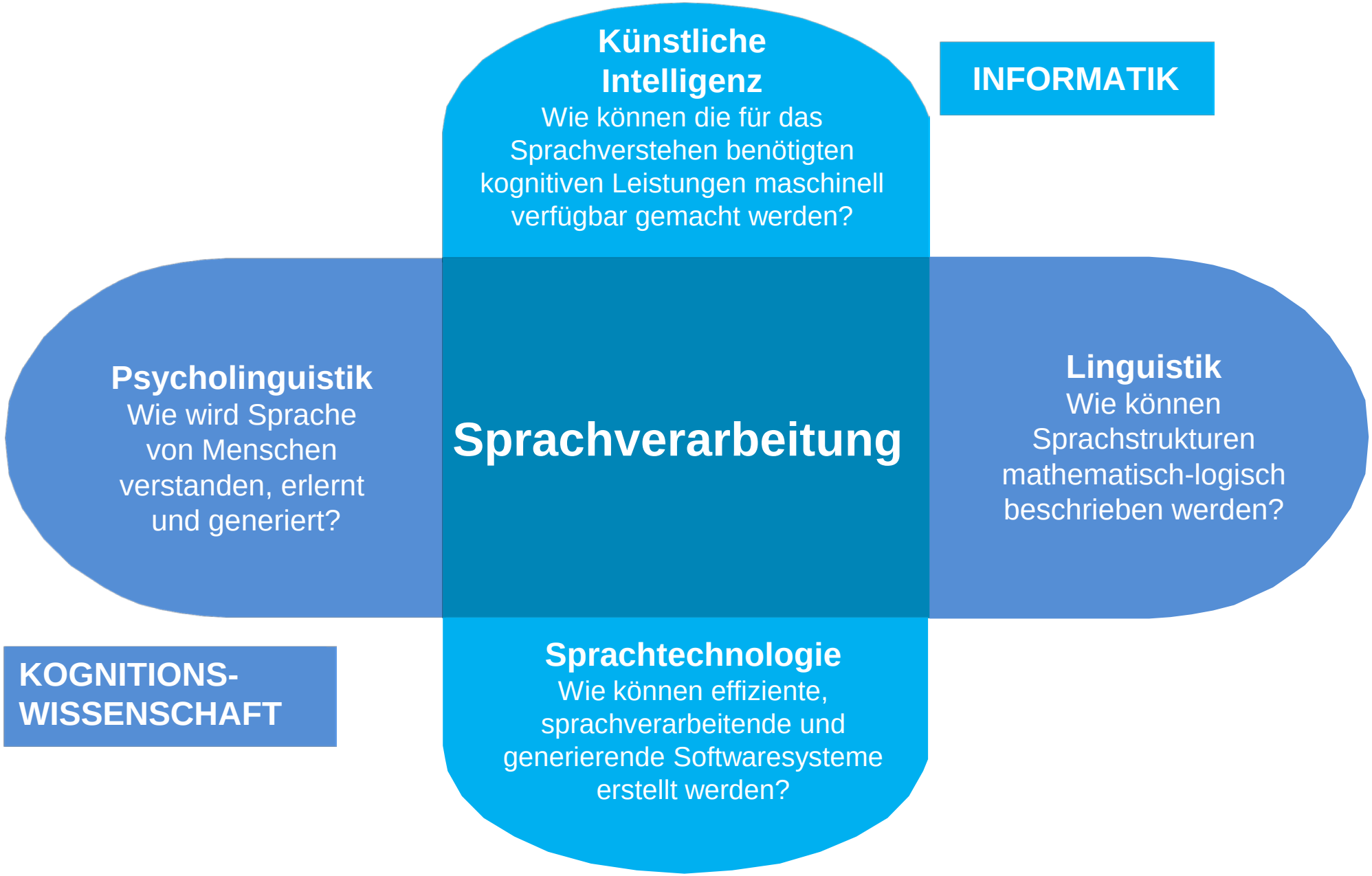
Sprachverhalten ist eine definierende Eigenschaft unserer Spezies

- **Können Computermodelle natürliches Sprachverhalten analysieren und simulieren?**

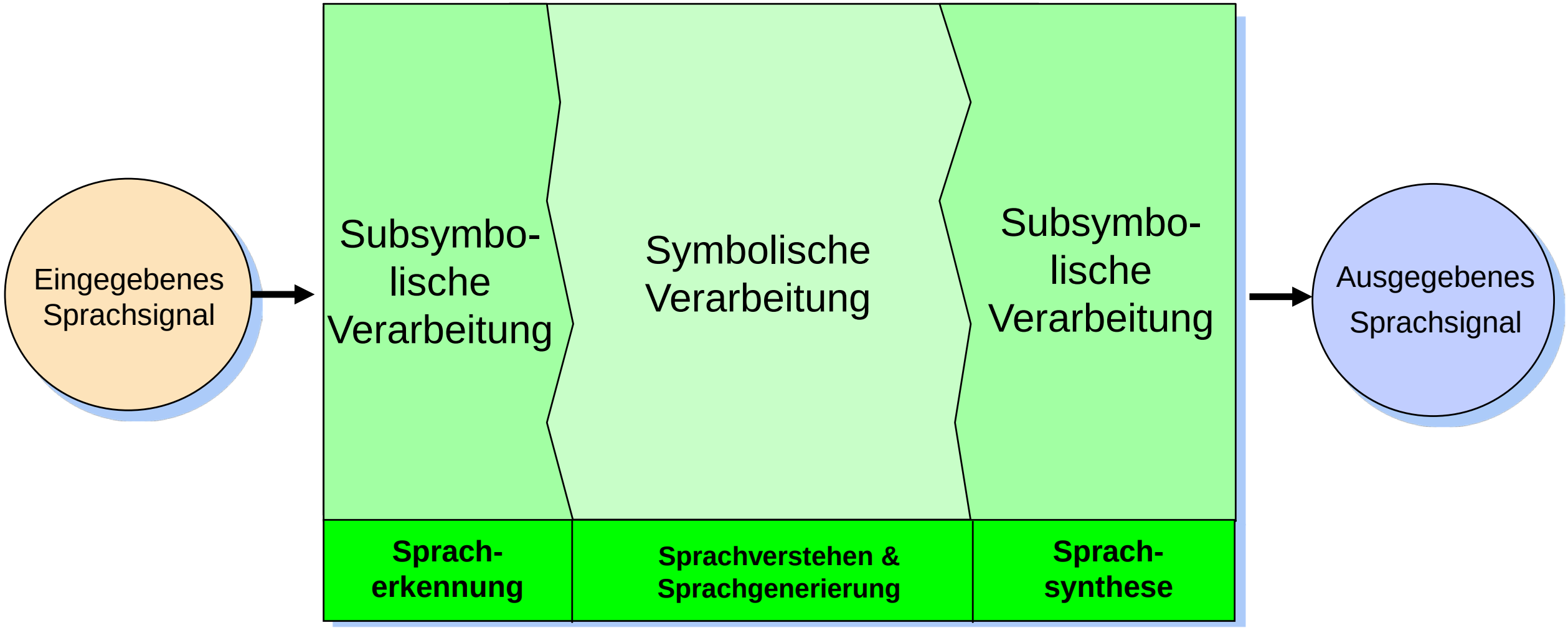
- **Können Computer zu echten Dialogpartnern werden?**

- **Kann der uralte Menscheitstraum, mit den Dingen sprechen zu können, mit Computern realisiert werden?**

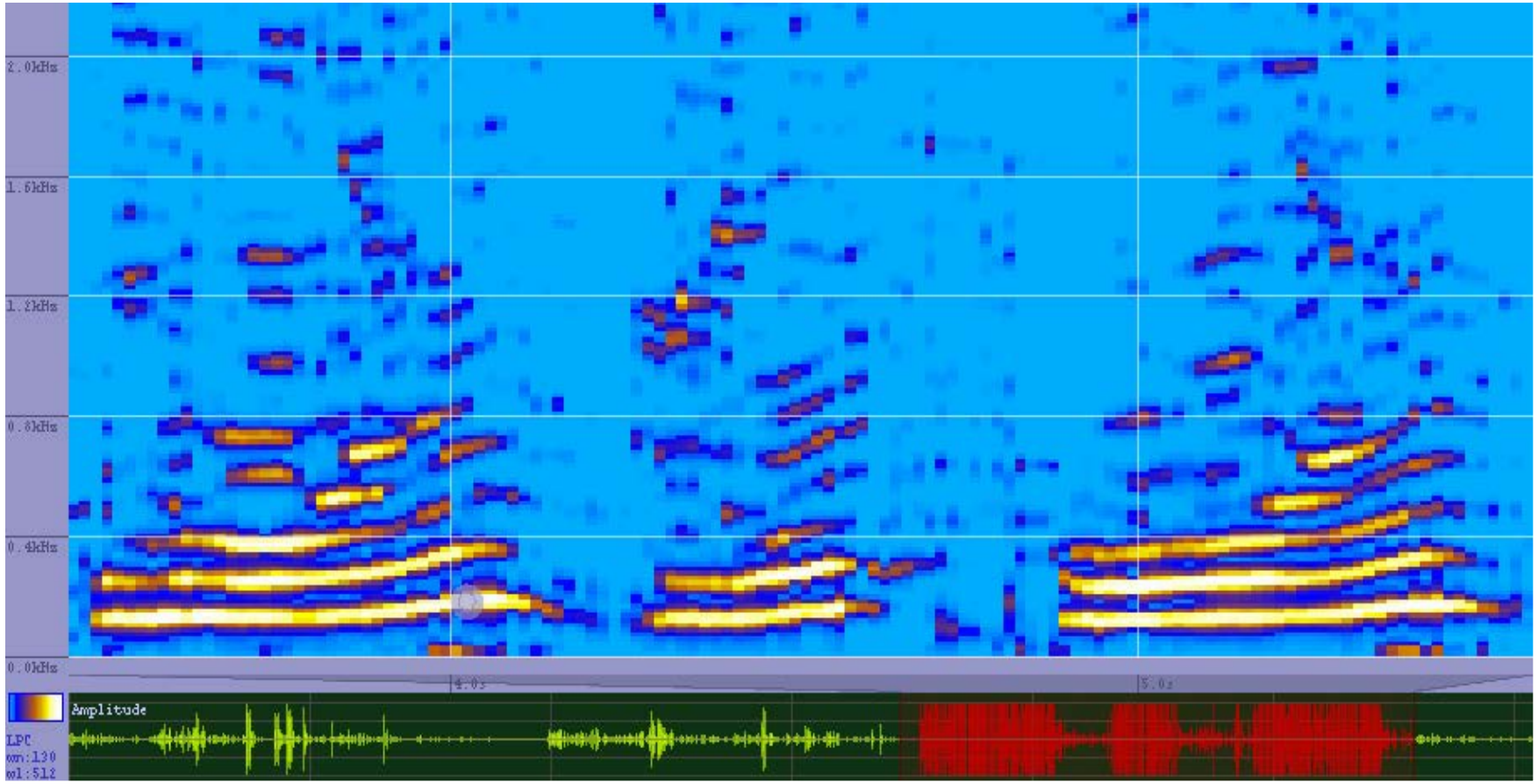
= eines der ehrgeizigsten wissenschaftlichen Ziele unseres Zeitalters


Künstliche Intelligenz
Wie können die für das Sprachverstehen benötigten kognitiven Leistungen maschinell verfügbar gemacht werden?
INFORMATIK
Psycholinguistik
Wie wird Sprache von Menschen verstanden, erlernt und generiert?
Sprachverarbeitung
Linguistik
Wie können Sprachstrukturen mathematisch-logisch beschrieben werden?
KOGNITIONS-WISSENSCHAFT
Sprachtechnologie
Wie können effiziente, sprachverarbeitende und generierende Softwaresysteme erstellt werden?

Signal-Symbol-Signal Transformation bei der Computermodellierung von Sprachdialogsystemen



Warum ist Sprachverstehen für den Computer so schwer?



Urlauber wollen wieder me:r ans me:r“

Warum ist Sprachverstehen für den Computer so schwer?

- ☞ **Gleiche Schallwellen werden je nach Kontext zu verschiedenen Wörtern**

  Beispiel: „Urlauber wollen wieder me:r ans me:r“

  →Urlauber wollen wieder mehr ans Meer.

- ☞ **Viele Menschen sprechen Dialekt**

  Beispiel: „Isch find das nätt“

  Bedeutung (1) Ich finde das nett.

  oder (2) Ich finde das nicht.

Warum ist Sprachverstehen für den Computer so schwer?

Johannes Gutenberg Stiftungsprofessur

- **Wortgrenzen gehen im Sprachfluß unter:**

  Beispiel: „amontag“ → „am Montag“

- **Der Mensch spricht „ohne Punkt und Komma“**

  Beispiel: „So machen wir das vielleicht klappt es“

Bedeutung (1) So machen wir das. Vielleicht klappt es.

oder (2) So machen wir das vielleicht. Klappt es?

Warum ist das Sprachverstehen für den Computer so schwer?

Johannes Gutenberg Stiftungsprofessur

- **Bei spontaner Rede entstehen viele Versprecher**

  Beispiel: „Wir treffen uns dann am Mon, äh, am Dienstag.“

- **Dialogpartner fallen dem Sprecher oft „ins Wort“**

Beispiel: System: „Können wir dann am | Mittwoch zum Essen...“

Sprecher: | „Da kann ich nicht.“

Warum ist das Sprachverstehen für den Computer so schwer?

Johannes Gutenberg Stiftungsprofessur

☞ **Der Redefluß leitet häufig in die Irre**

Beispiel: „Die Staatssekretärin lobt... der Ministerpräsident.“

~~**Subjekt: Staatssekretärin Prädikat: lobt Objekt: ??**~~

Subjekt: Ministerpräsident Prädikat: lobt Objekt: Staatssekretärin

☞ **Viele Formulierungen sind mehrdeutig**

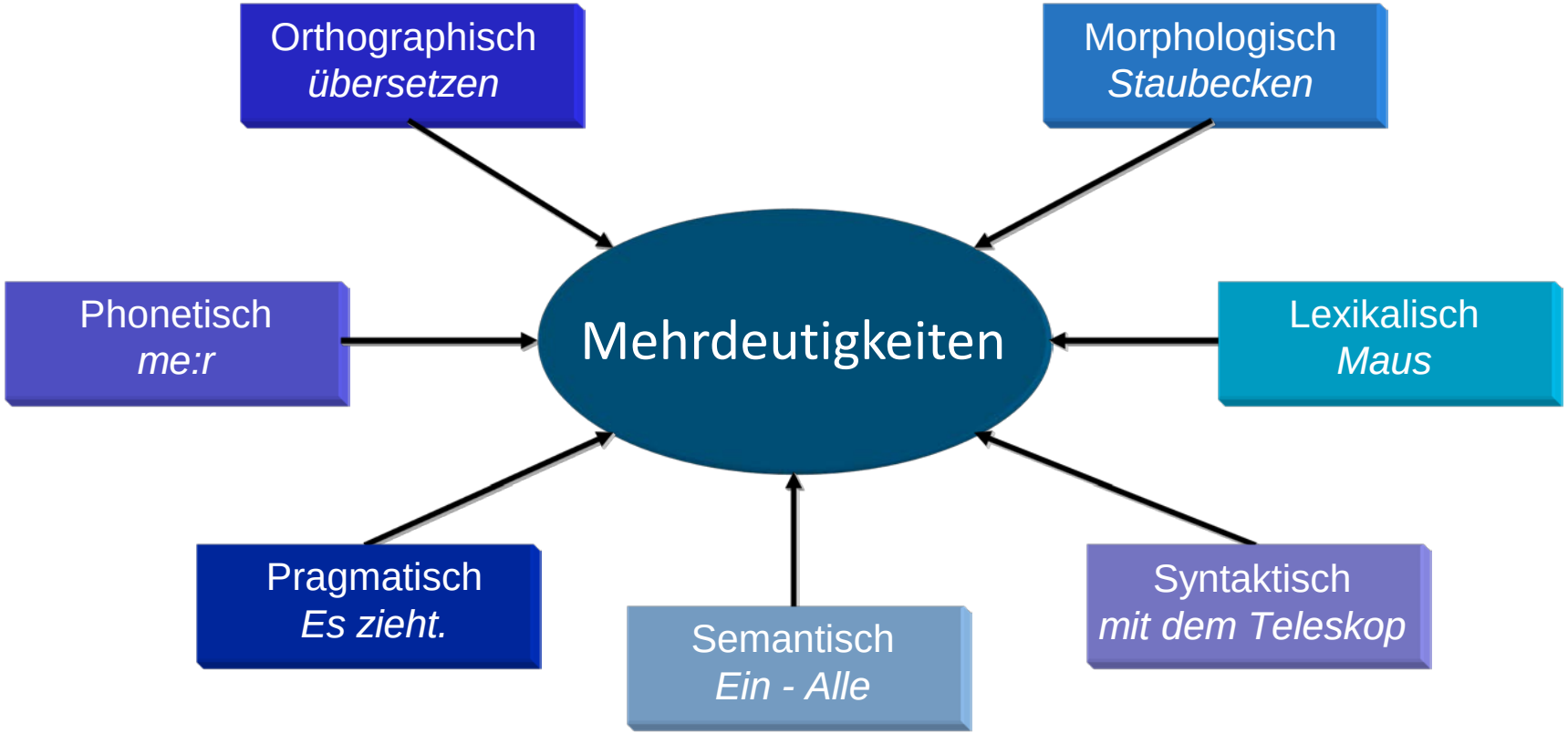
Beispiel: „Wir telefonierten mit Freunden in Japan.“

Bedeutung (1) Wir telefonierten (mit Freunden in Japan).

oder (2) (Wir telefonierten mit Freunden) in Japan.

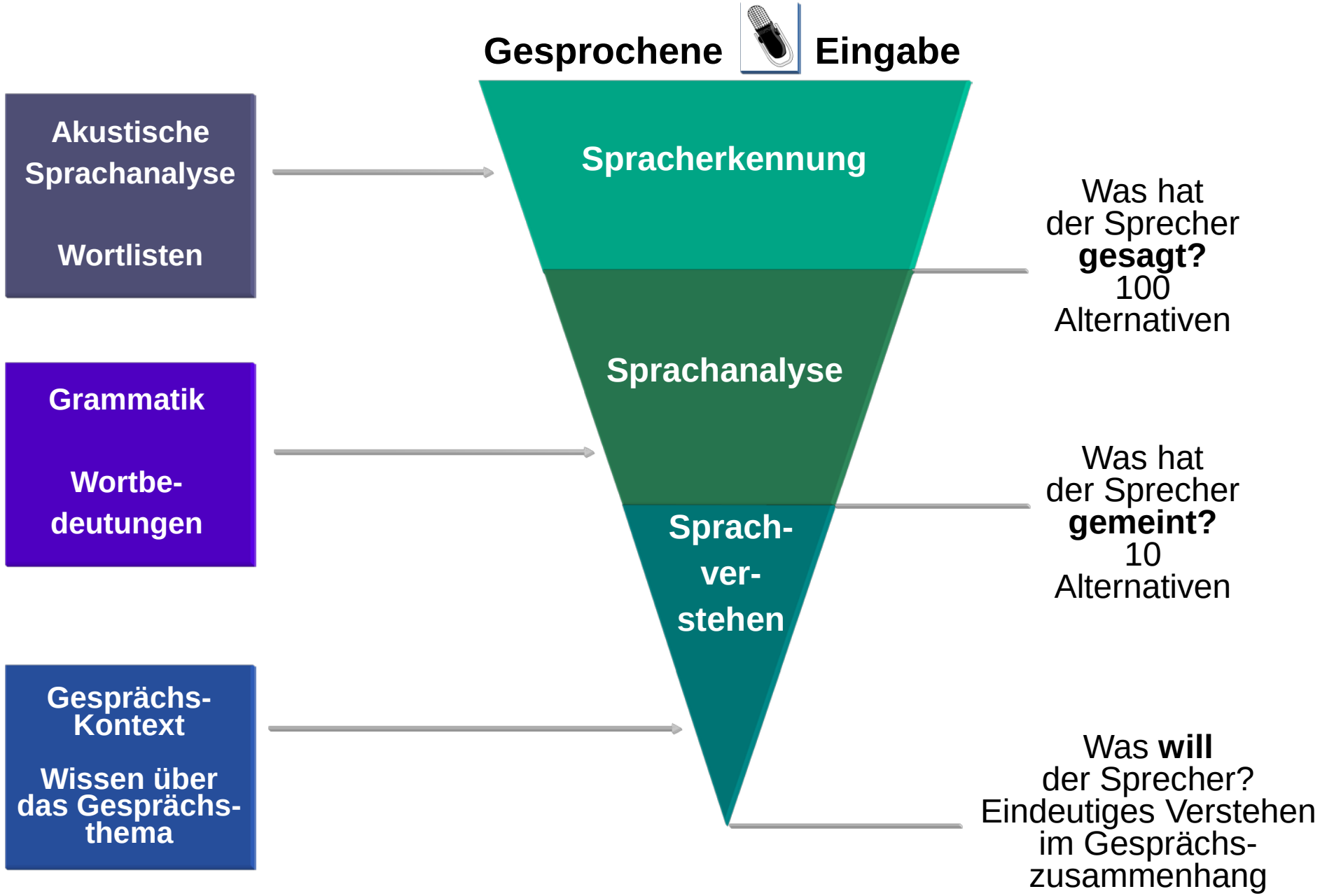
Auflösung mehrdeutiger sprachlicher Äußerungen

- **Problem der kombinatorischen Explosion der Lesarten durch Propagierung von Alternativen über alle Verarbeitungsebenen**
- **Durch die Unsicherheit bei der Spracherkennung entstehen Wörtergitter mit alternativen Hypothesen, welche die Flut von Lesarten noch weiter erhöhen**



Drei Stufen der Sprachverarbeitung

Johannes Gutenberg Stiftungsprofessur



Von der Eingabeschallwelle zur Ausgabeschallwelle

Verbmobil: Dialogdolmetschen von Spontansprache in der Reisedomäne

Automatisches Verstehen und Korrigieren von Versprechern in spontanen Telefondialogen

Wir treffen uns in Mannheim, äh, in Saarbrücken.

Deutsch

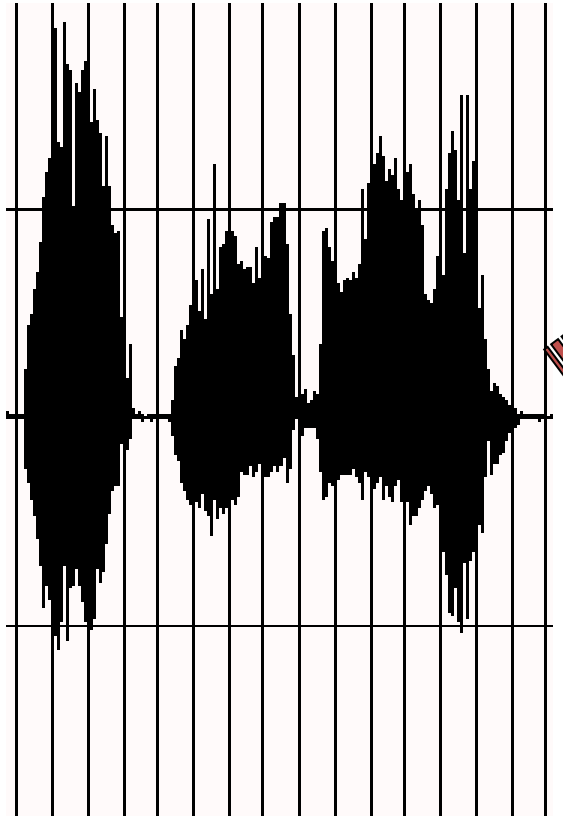
Englisch

We are meeting in Saarbruecken.

Spracherkennung: Vom Signal zur Worthypothese

Johannes Gutenberg Stiftungsprofessur

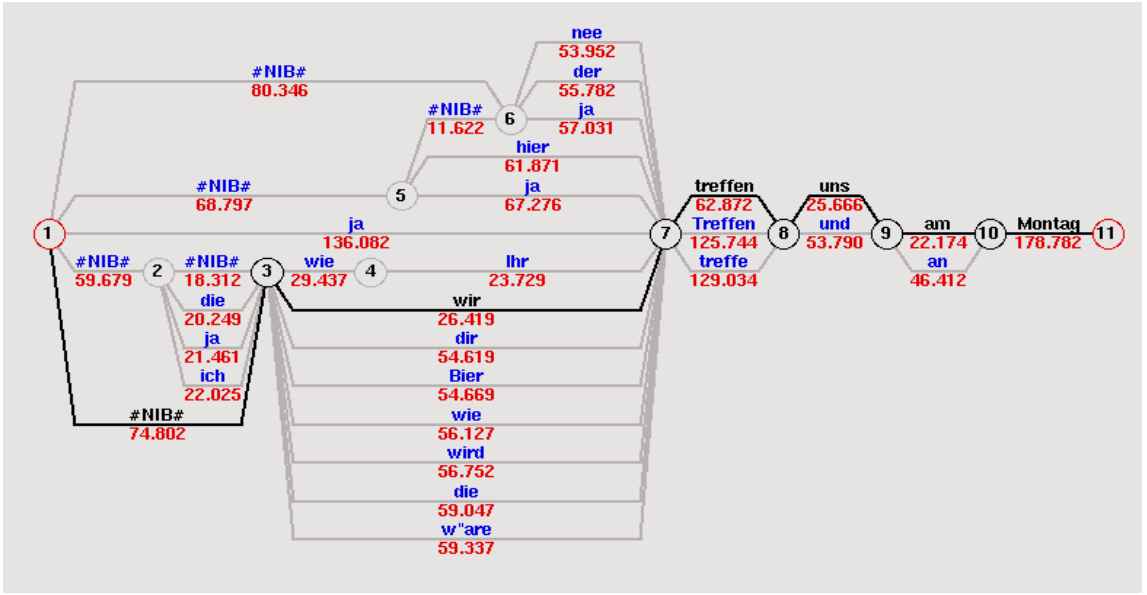
Audiosignal



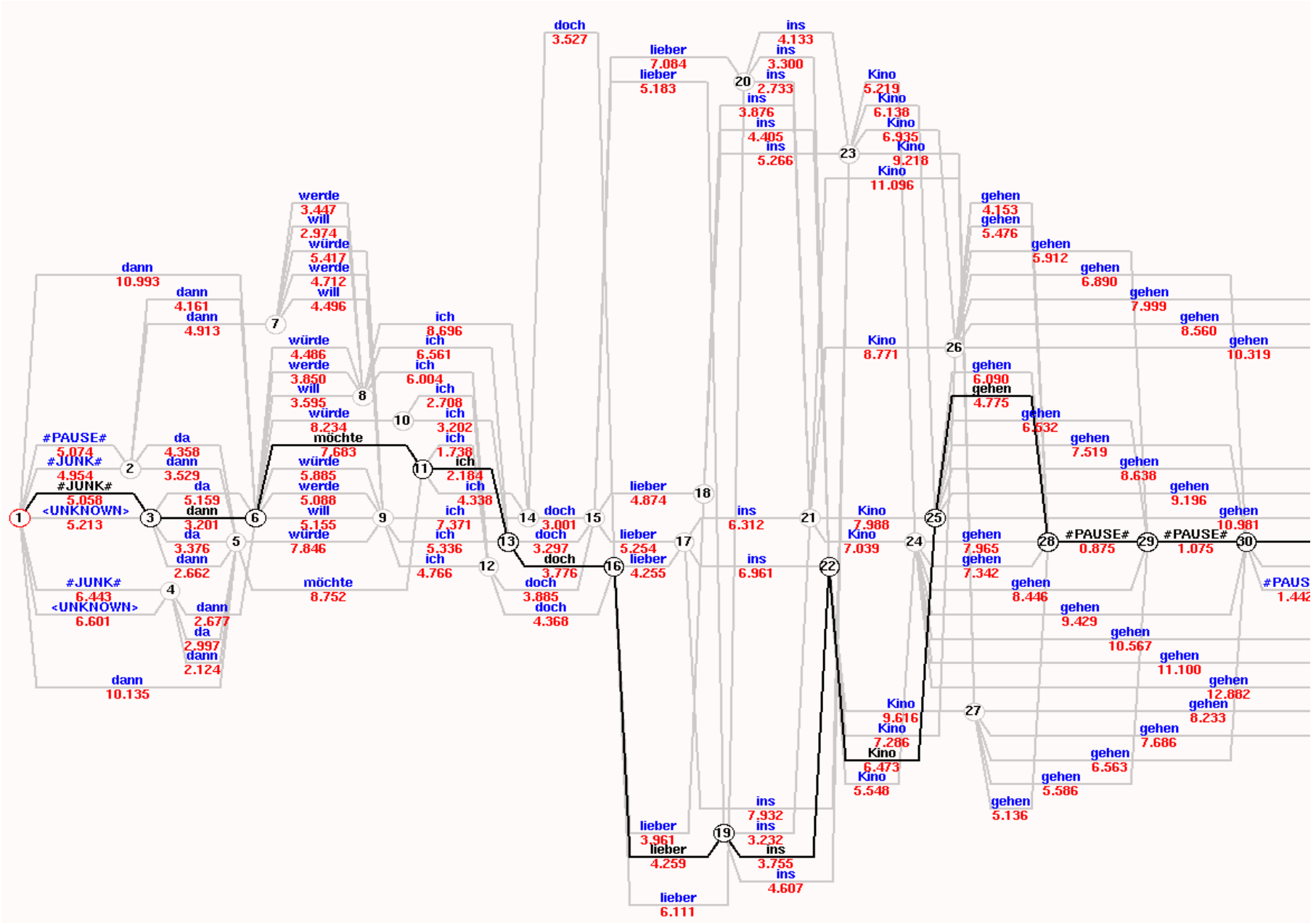
Sprachidentifikation und -erkennung

- Deutsch
- Englisch
- Japanisch
- Chinesisch

Worthypothesengraph



Unsicheres Ergebnis der Spracherkennung



Weltwissen im Computermodell für die maschinelle Bedeutungsanalyse

Vater zu einem Service-Roboter im Cyber-Restaurant:

(1) Die Apfelschorle trinkt meine Tochter, die Weinschorle meine Frau.

(A) trinkt (Agens: Apfelschorle, Objekt: Tochter) ∧
trinkt (Agens: Weinschorle, Objekt: Frau)

Weltwissen: Apfelschorle, Weinschorle ⊑ Getränk
Tochter, Frau ⊑ Mensch

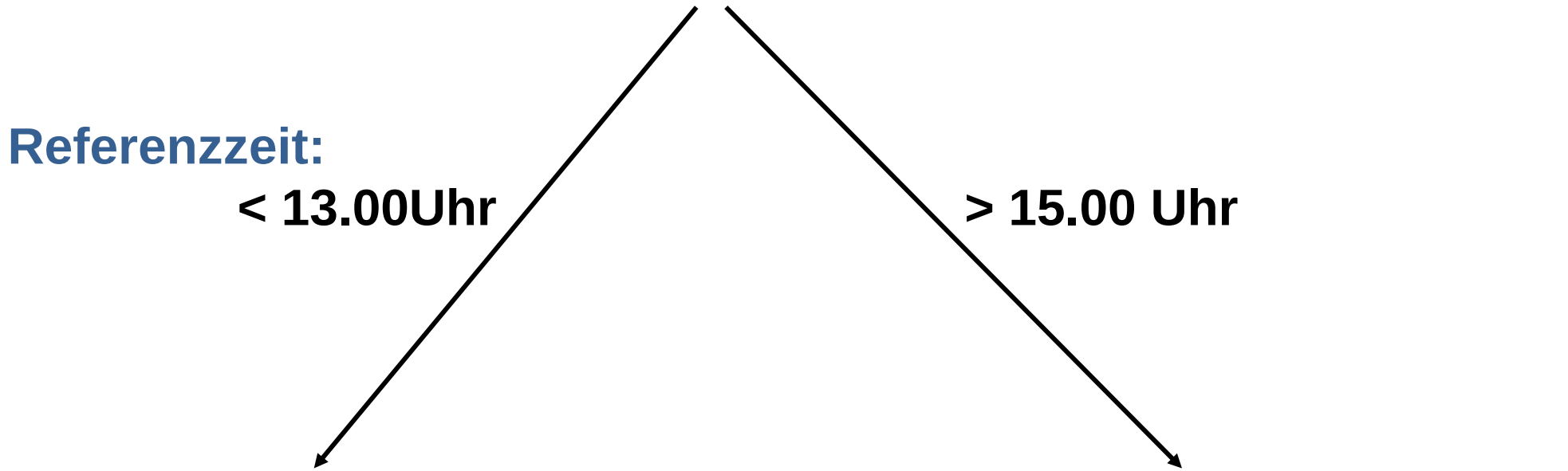
Selektionsrestriktion: trinkt (Agens: Mensch, Objekt: Getränk)

(B) trinkt (Agens: Tochter, Objekt: Apfelschorle) ∧
trinkt (Agens: Frau, Objekt: Weinschorle)

Semantische Präzisierung durch Situationswissen

Johannes Gutenberg Stiftungsprofessur

Lassen Sie uns zusammen Essen gehen!



Let‘s have lunch together!

Let‘s have dinner together!

Berücksichtigung des Kontextes und von Weltwissen beim Semantischen Transfer

Johannes Gutenberg Stiftungsprofessur

Beispiel: Platz ➔ room / table / seat

1 Nehmen wir dieses Hotel, ja. ➔ Let us take this hotel.

Ich reserviere einen **Platz**. ➔ I will reserve a **room**.

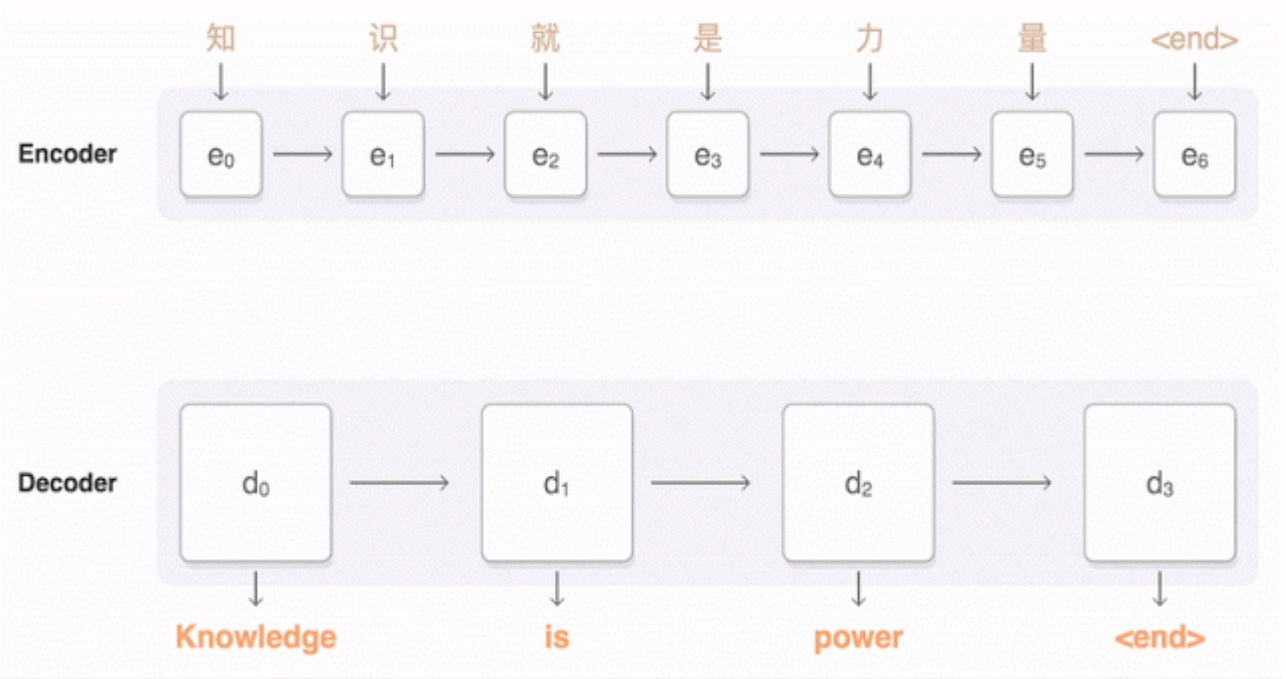
2 Machen wir das Abendessen dort. ➔ Let us have dinner there.

Ich reserviere einen **Platz**. ➔ I will reserve a **table**.

3 Gehen wir ins Theater. ➔ Let us go to the theater.

Ich möchte **Plätze** reservieren. ➔ I would like to reserve **seats**.

Ende-zu-Ende-Lernen für die neuronale Übersetzung: Chinesisch-Englisch (Google GNMT)



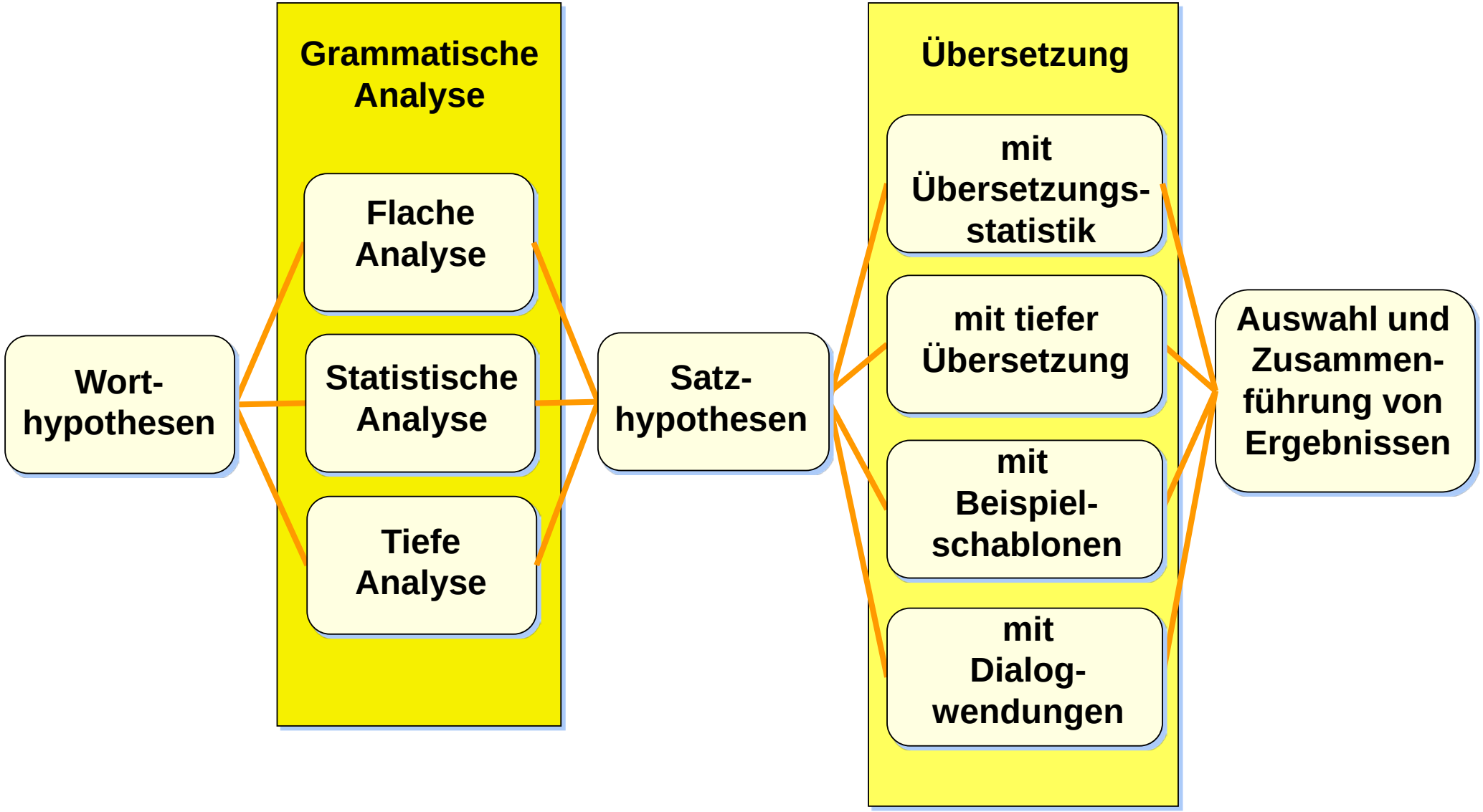
NMT analysiert die Eingabe als Ganzes und behält alle Wörter und Phrasen im Blick.

Dazu kodiert das Übersetzungssystem zuerst die chinesischen Wörter in eine Liste von Vektoren, in der jeder Vektor die Bedeutung aller bis dahin gelesenen Wörter repräsentiert (Encoder).

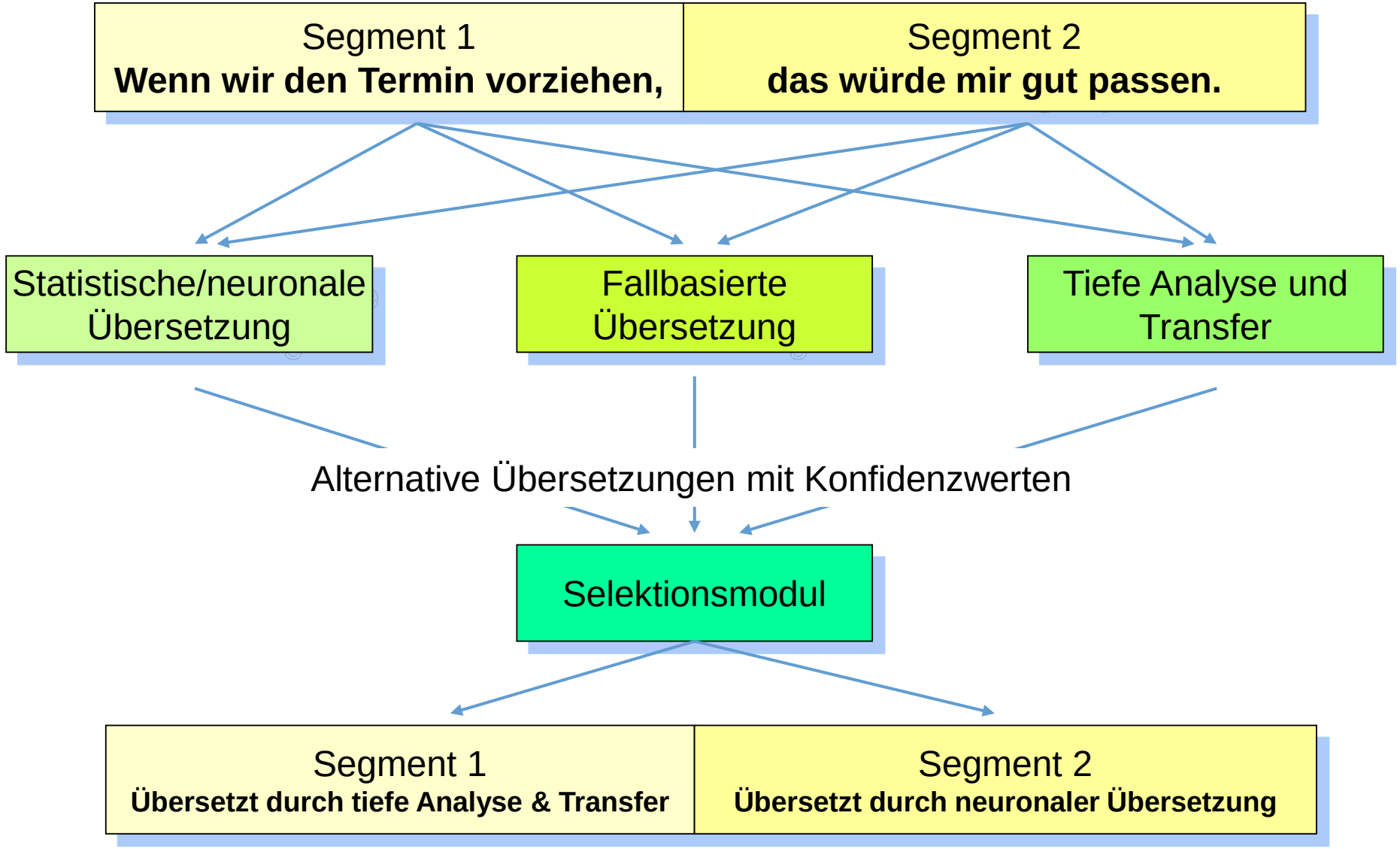
Danach wird der englische Satz wortweise generiert (Decoder).

Dabei achtet der Decoder auf eine gewichtete Verteilung der kodierten chinesischen Vektoren, die am wichtigsten zur Generierung der englischen Entsprechung sind (Attention).

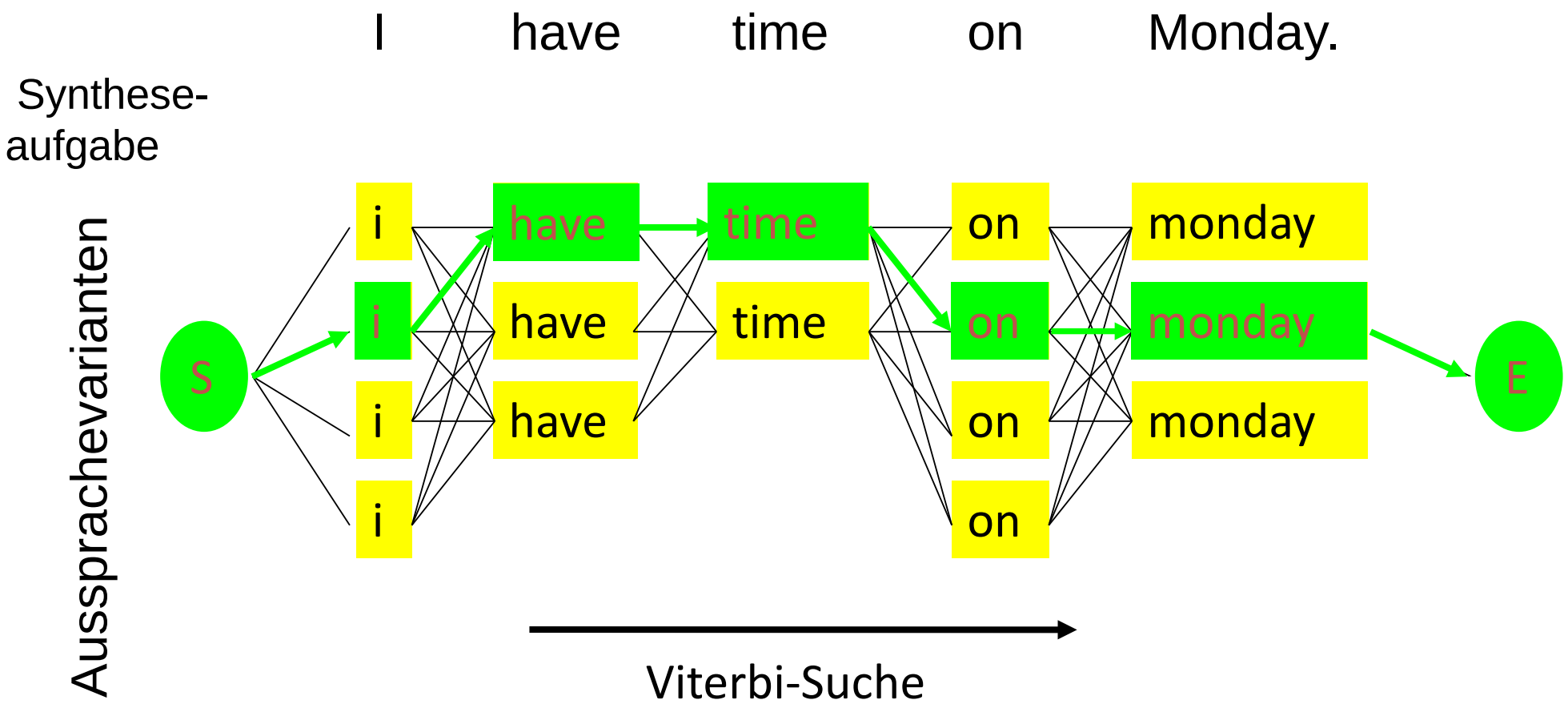
Zusammenführung von flacher und tiefer Sprachverarbeitung in Verbmobil



Die Kombination von Segmentübersetzungen aus einem Mix an Übersetzungsverfahren



Corpus-basierte Sprachsynthese



Dynamische Auswahl von Aussprachevarianten für Wörter je nach Funktion im Satz

Anthropomorphe Objekte: Sprechende Produkte

Durch Künstliche Intelligenz mit allen Sinnen ins Internet

Sprache

Blickbewegung

Gestik

Biometrie

Multimodale Interaktion

Mimik

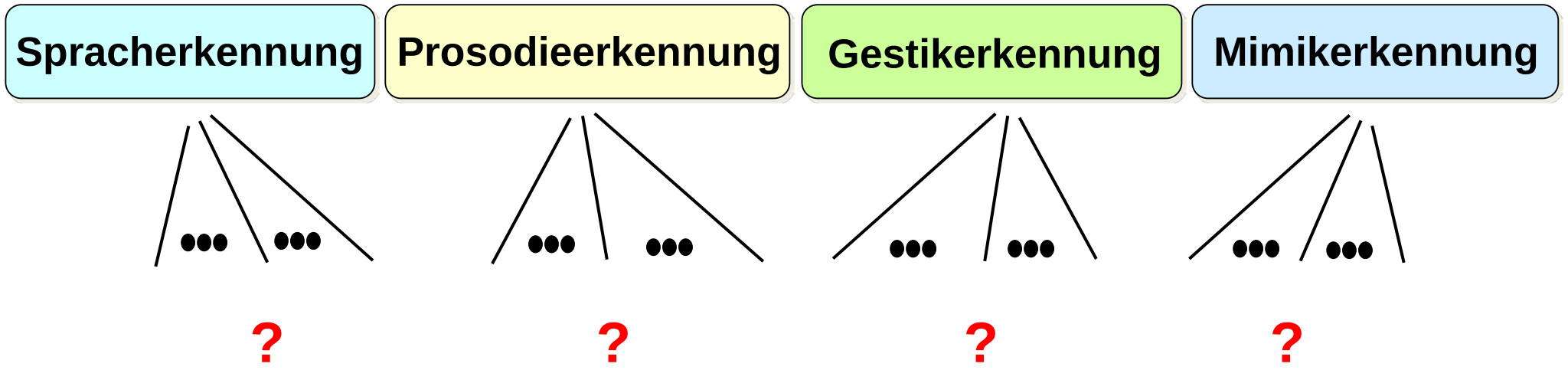
Physische Aktion

Körperhaltung

Die Fusion multimodaler Eingaben

Johannes Gutenberg Stiftungsprofessur

Multiple Modalitäten erhöhen die Interpretationsunsicherheit

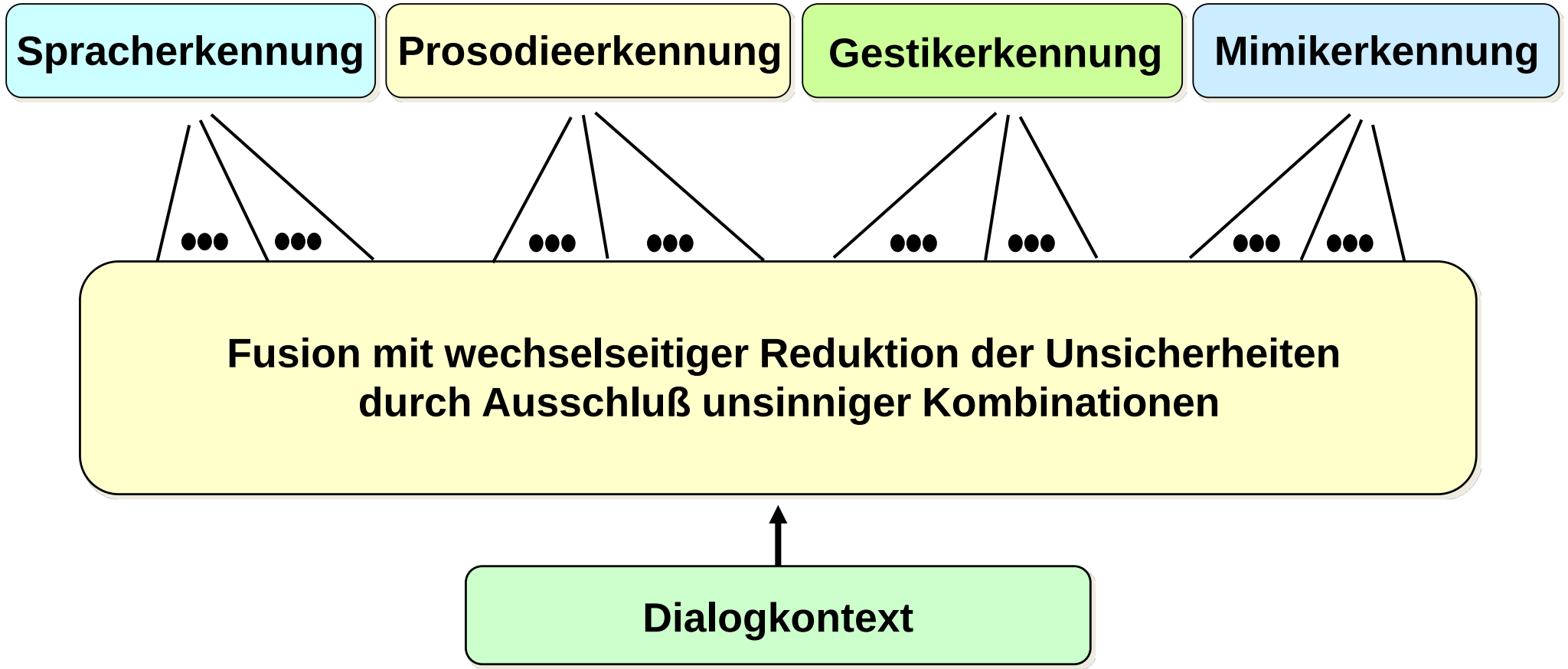


Unsicherheiten bei der Signalinterpretation in perzeptiven Benutzerschnittstellen

Die Fusion multimodaler Eingaben

Multiple Modalitäten erhöhen die Interpretationsunsicherheit

aber: die semantische Fusion multipler Modalitäten macht die Interpretation im Kontext eindeutig



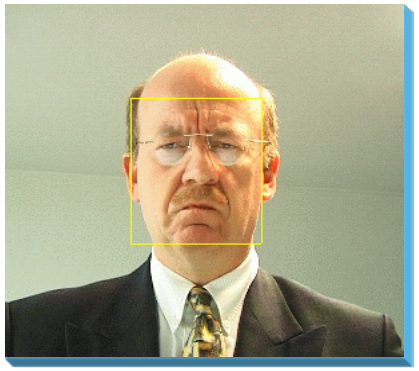
Fusion von Sprach- und Mimikerkennung

Johannes Gutenberg Stiftungsprofessur

Modifikation der Standardsemantik (Ironie, Sarkasmus)

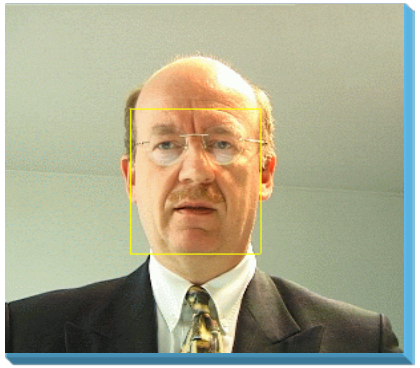
(1) System: : Hier sehen Sie die Übersicht zum heutigen ZDF-Programm.

(2) Benutzer: Echt toll.



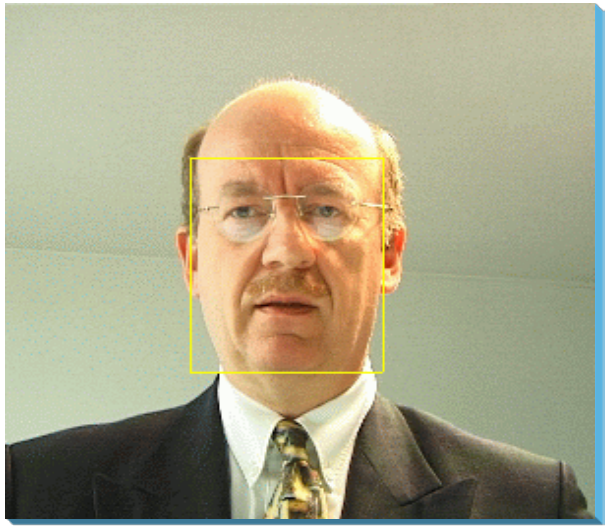
(3) System: Ich zeige Ihnen alternativ das Programm eines anderen Senders.

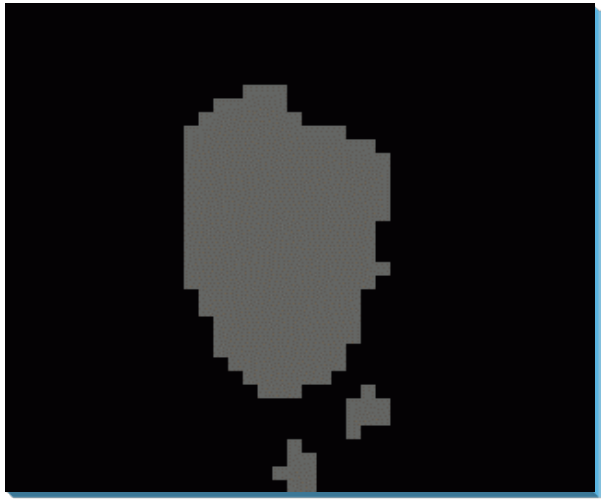
(2') Benutzer: Echt toll.



(3') System: Welche Sendungen wollen Sie aus dem ZDF-Programm sehen oder aufzeichnen?

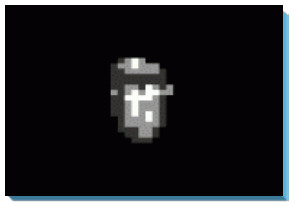
Videobasierte Mimikerkennung auf der Basis von Eigenfaces



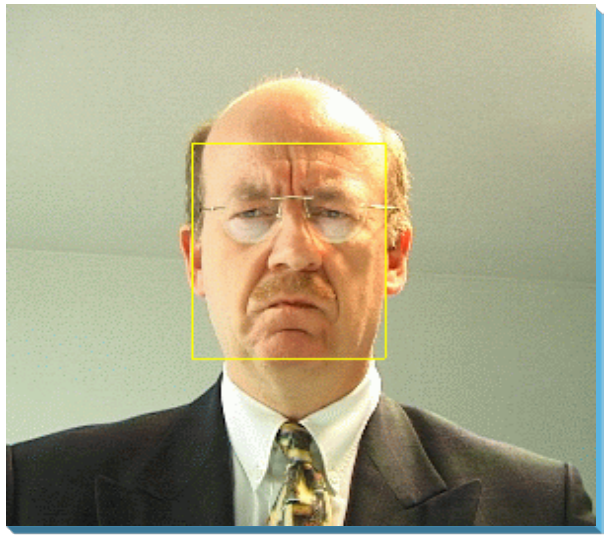


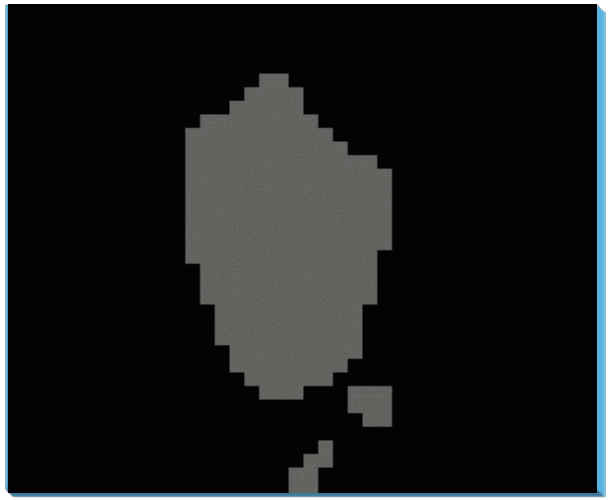
neutral

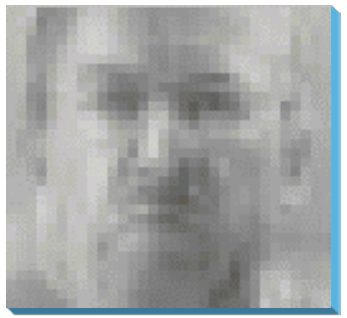
ärgerlich



Sprecherunabhängie Emotionserkennung

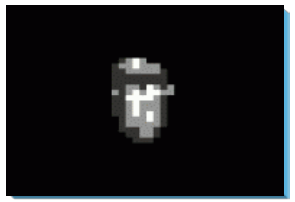






neutral

ärgerlich



Kombination von Sprache und Aktionskontext

Interpretation und Disambiguierung von multimodalen Äußerungen: Auswahl der **Scheibenwischer**geschwindigkeit vs. wechseln der **Tempomat**einstellung

Multimodale Fahrereingaben: Sprache und Blickbewegung

Korrekturdialog während der Fahrt in unserem Twitter4Car in unserem DFG-Exzellenzcluster Multimodal Computing and Interaction

TALK: Robuste Semantische Sprachdialogverarbeitung im BMW

Sprachdialog mit einem iPhone in TEXO-Theseus zur Selektion von Webdiensten

Künstliche Intelligenz für unterwegs: Persönliche digitale Assistenten auf SmartPhones

Apple: Siri

Google: Assistant

Microsoft: Cortana

Samsung: Viv

2007 in Deutschland: Siri-Urversion SmartWeb vom DFKI mit Telekom und Siemens (Ende Handy-Sparte 2005)

Hololens 2016 für freihändige Assistenzfunktionen

Blindtest unserer Antwortmaschine SmartWeb als Super-Siri

Johannes Gutenberg Stiftungsprofessur

Noch nie gehörte Frage: „Wer hat bei den Salzburger Festspielen letztes Jahr die Premiere in La Traviata gesungen?“ wird von SmartWeb verstanden und eine Antwort für 2007 wird aus Dokumenten im Internet gefunden.

Künstliche Intelligenz für die gesamte Familie zuhause

Amazon: Echo - Alexa

Google: Home

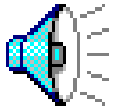
Jibo (mit Kamera)

- Funktionen: Steuerung der SmartHome-Funktionen, Einkaufslisten, Bestellungen, Zugriff auf Wissens-bank mit Nachfragen, Terminverwaltung, Musik- und Videosuche
- 7 Offene **Fernfeldmikrophone** (abschaltbar), KI-Funktionen über **Cloud-Zugriff**
- Adaption durch **maschinelles Lernen** an die einzelnen Familienmitglieder

VirtualHuman: Mehrpersonenkommunikation mit komplexen Rednerwechselstrategien

QRIO: Sprachdialoge eines Unterhaltungsroboters mit einer Kindergruppe

Sprachbasierte Adaption an die Altersklasse des Benutzers

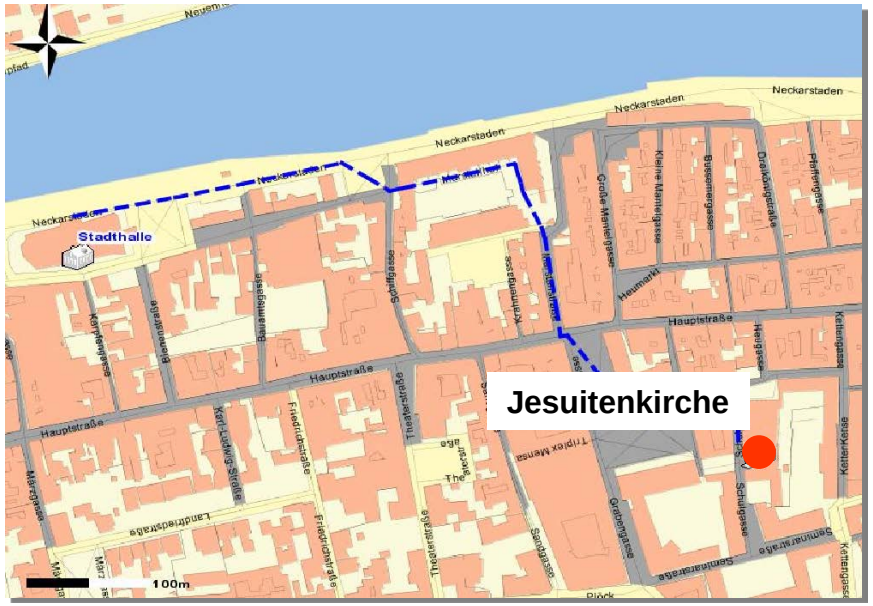


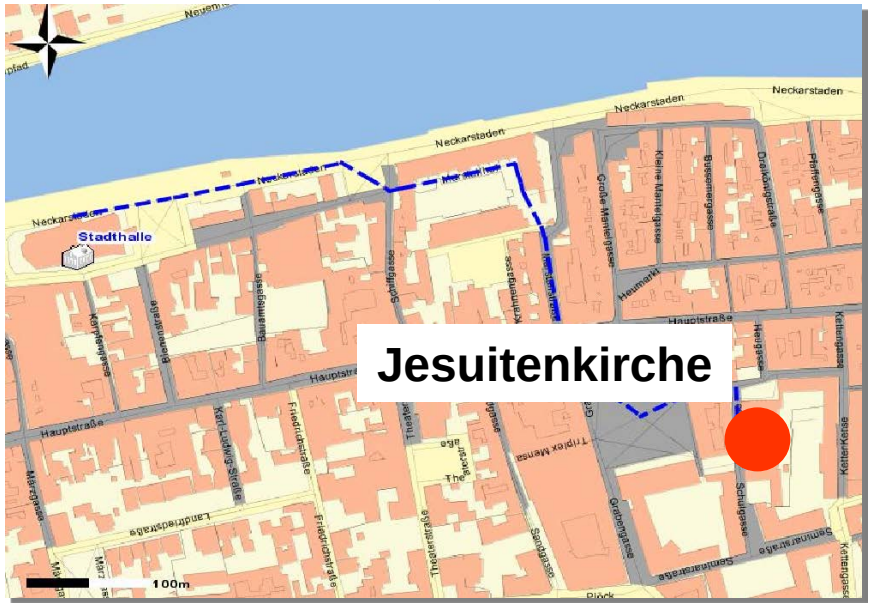
Versteifung des Stimmlippengewebes im Alter

Jitter und Shimmer

Erkennung des biologischen Alters der Stimme durch akustische Merkmale

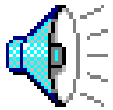
Jitter und Shimmer





Sprachbasierte Adaption an die Altersklasse des Benutzers

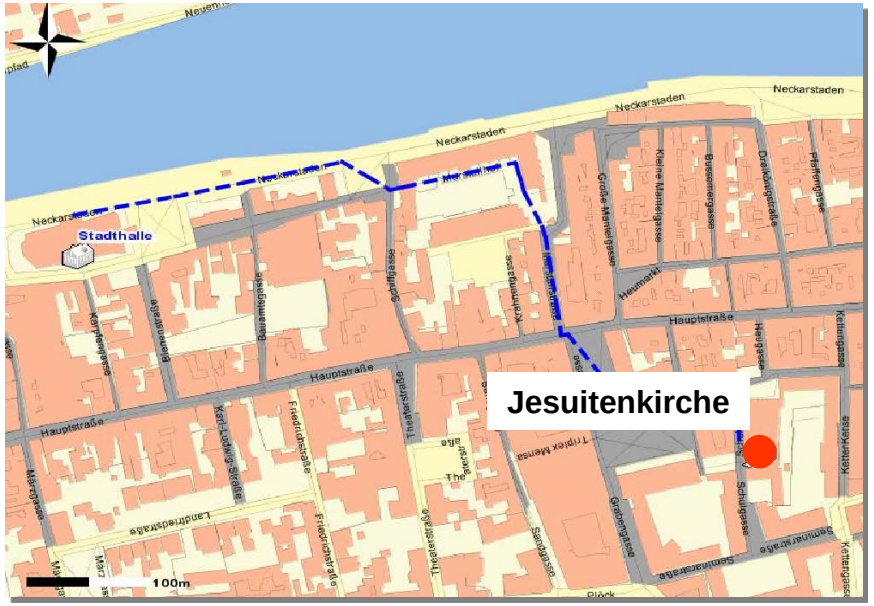
Versteifung des Stimmlippengewebes im Alter

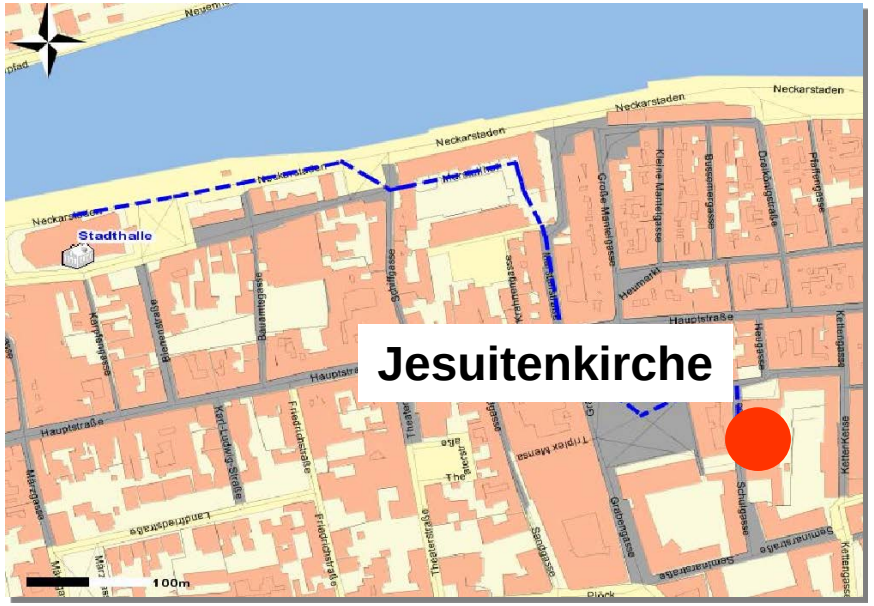


Jitter und Shimmer

Erkennung des biologischen Alters der Stimme durch akustische Merkmale

Jitter und Shimmer





Auswirkungen des Alterns auf die Sprachproduktion:

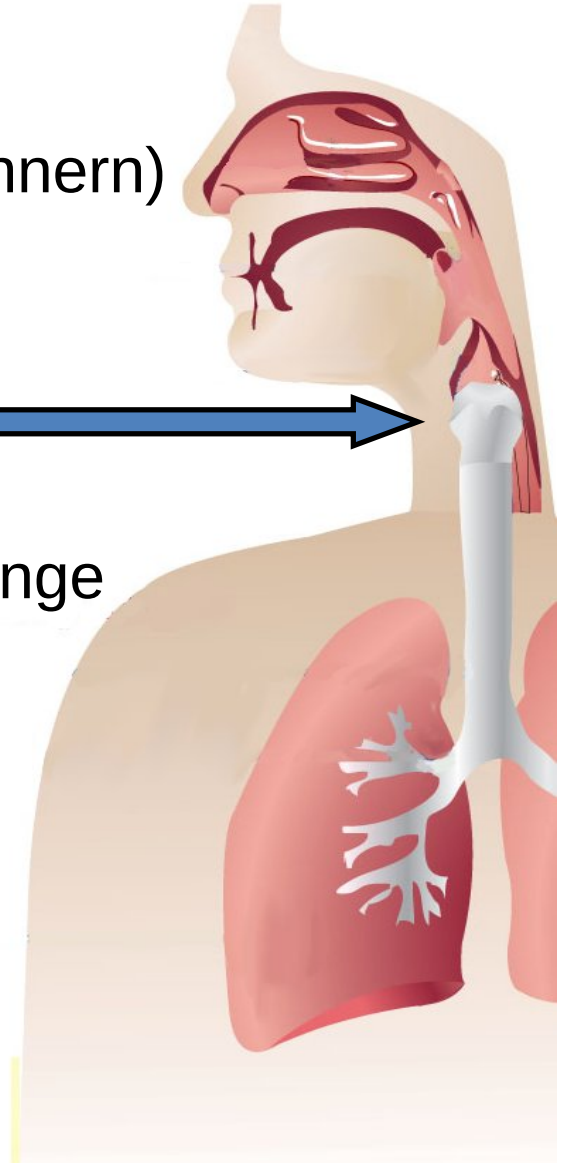
Johannes Gutenberg Stiftungsprofessur

zu erwartende Folgen:

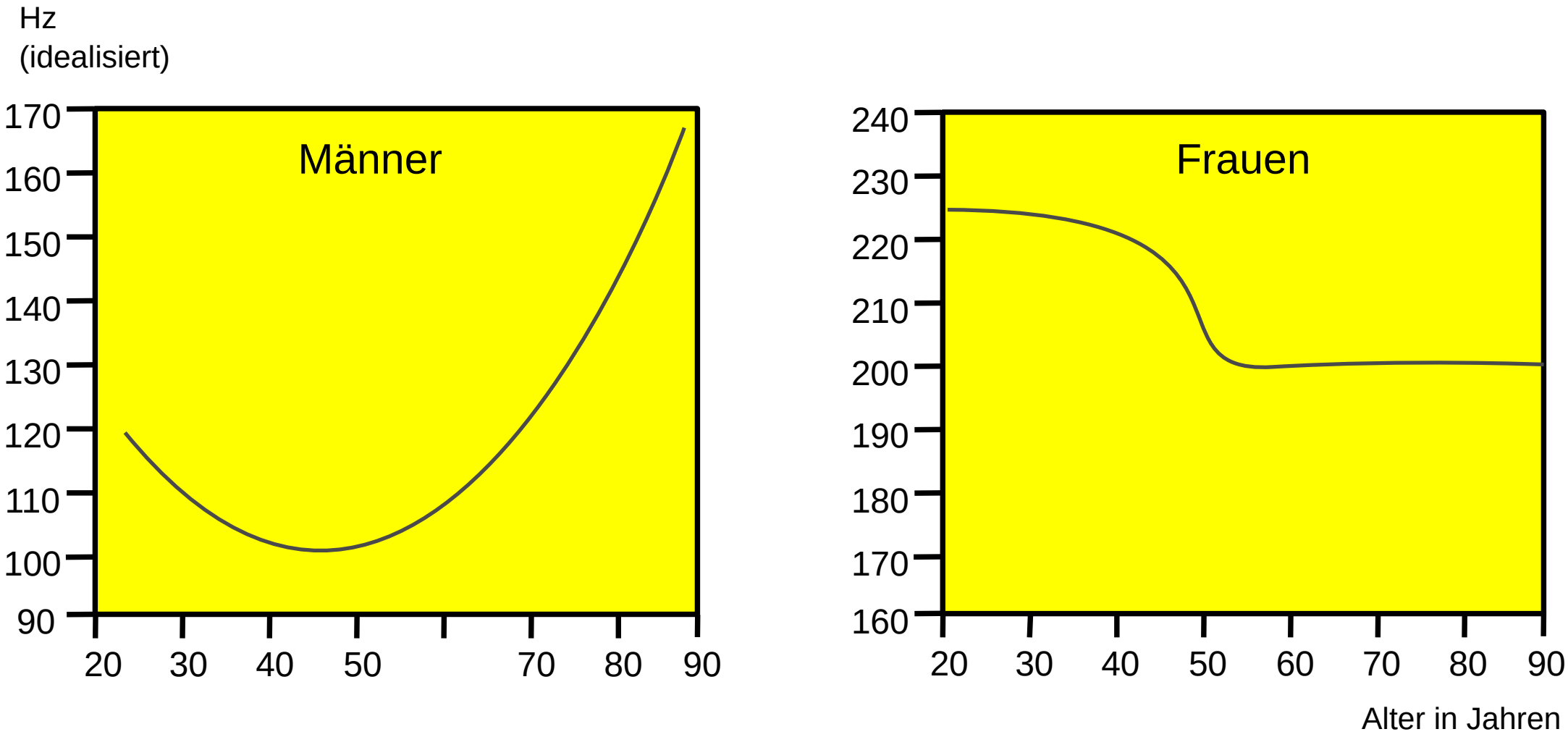
Ansteigen der Grundfrequenz F0 (bei Männern)

Verminderung der Stimmqualität

Kehlkopf	Verkalkungs- und Verknöcherungsvorgänge
Stimmlippen	Versteifung Gewebeschwund



Entwicklung der Grundfrequenz der Stimme im Alter



Automatische Anpassung des Dialogverhaltens durch biometrische Sprecherklassifikation

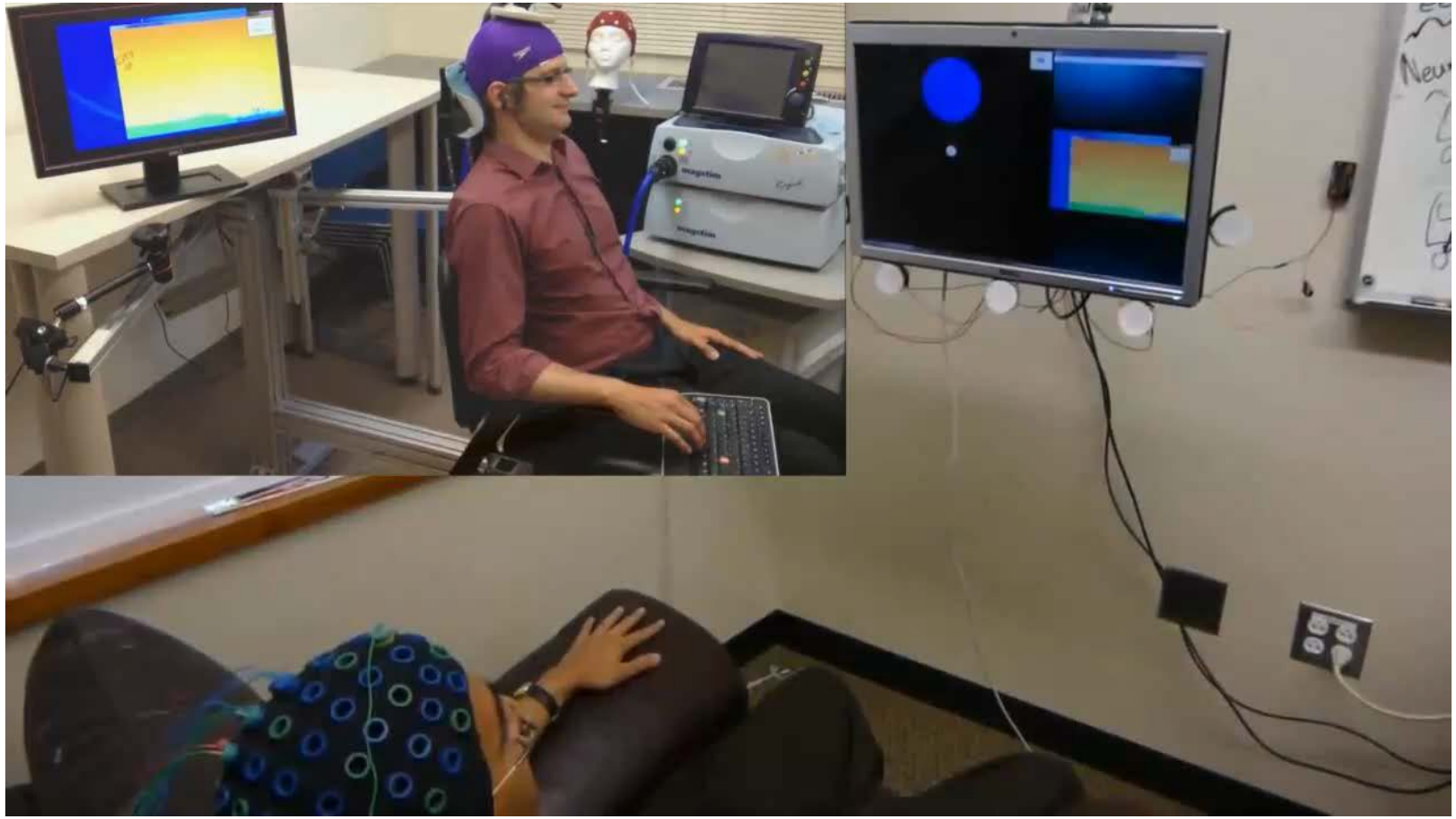
Johannes Gutenberg Stiftungsprofessur

Kooperation der T-Labs der Deutschen Telekom und DFKI

Internet-Uhr zur Innenraumnavigation sowie Fahrstuhlsteuerung des DFKI - M2M Kommunikation und Spracheingabe für Endbenutzer

Skype mit Simultanübersetzung à la Verbmobil des DFKI

Gehirn-zu-Gehirn-Kommunikation



Gedanke an Tastendruck des Senders (EEG)
löst diesen unwillentlich beim Empfänger aus (TMS)

Sprachverstehen und –generierung mit Künstlicher Intelligenz ist heute möglich, wenn auch noch nicht mit Qualität menschlicher Dialogpartner

In den 70iger Jahren noch reine Grundlagenforschung, heute gibt eine Sprachindustrie mit über 100 Spezialunternehmen in Deutschland und z.B. über 200 Sprachdialogsysteme-men, die in Deutschland über 500 000 Gespräche am Tag automatisiert fallabschliessend führen.

Beispiele für erfolgreiche Anwendungen sind:

- **Automatisierte Bestellung, Reservierung und Auskunft**
- **Steuerung von Komfortfunktionen im Auto**
- **Sprachliche Programmierung von Robotern/Rollstühlen/Geräten**
- **Sprachanalyse im Bereich Sicherheit/Terrorismusbekämpfung**
- **Übersetzen und Dolmetschen für Informationsaustausch**

Wettbewerb 2017: Professionelle Übersetzer vs. Google Translate und Systran – 15:25 Sieg über KI

In 50 Minuten mussten Texte zwischen Koreanisch und Englisch übersetzt werden.

2 Übersetzer bewerteten die Ergebnisse mit 30 als max. Punktzahl

Mangelndes globales Kontextverständnis der neuronalen und statistischen Übersetzung

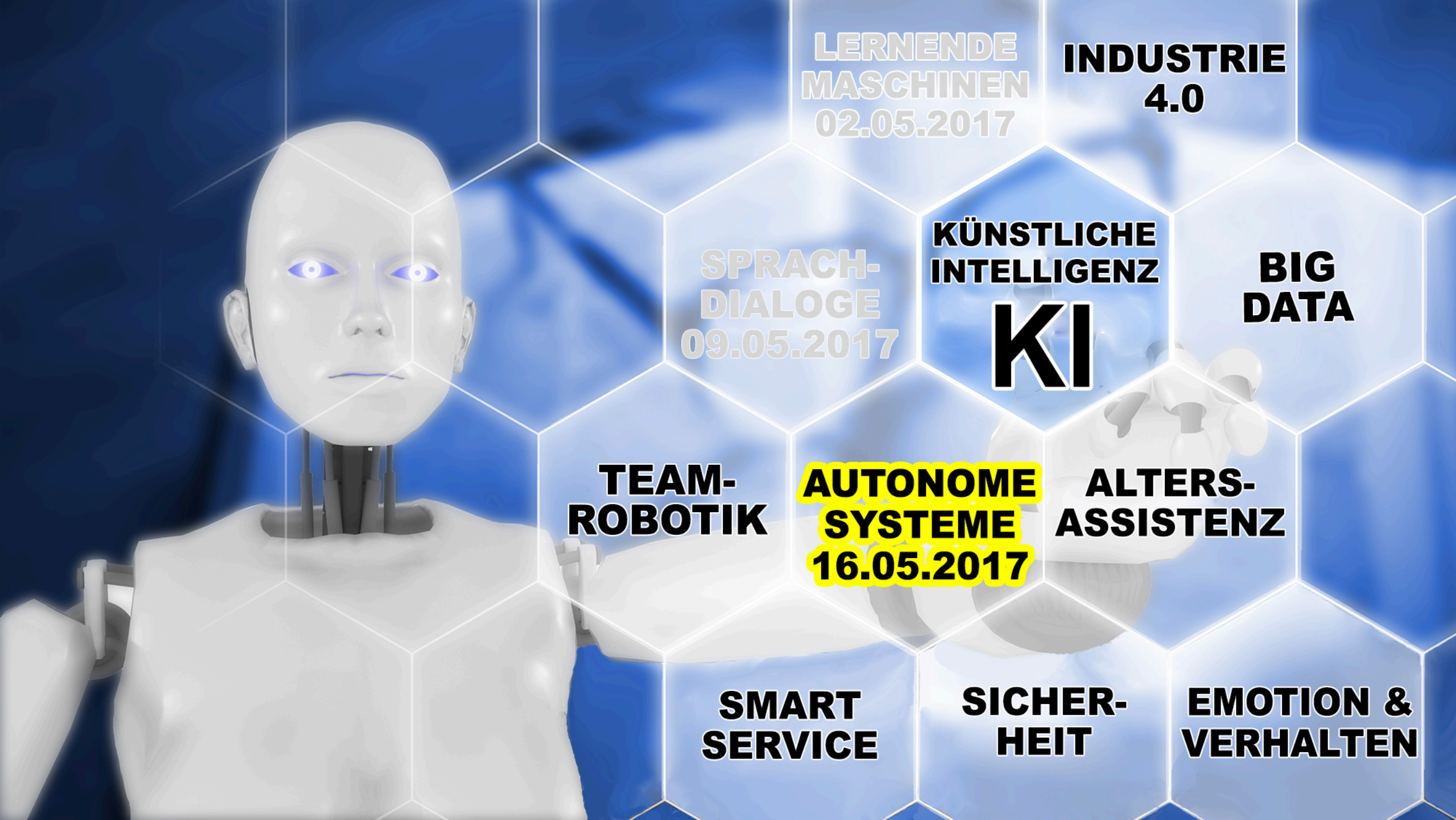
Heutige Grenzen beim KI-Sprachverstehen

Johannes Gutenberg Stiftungsprofessur

1. **Fehlende allgemeine Computermodelle für die Wechselwirkungen zwischen Sprachverstehen, Bildverstehen, Inferenz, Aktions-planung und Gedächtnis sowie anderen kognitiven Fähigkeiten.**

2. **Tiefes und robuste Verstehen derzeit nur für enge Themenbereiche, offene Gesprächsthemen erfordern Kombination von flachen und tiefen Verfahren**

3. **Maschinelles Lernen noch nicht für komplexe Dialoge und Spontansprache z.B. Verstehen von Wortschöpfungen aufgrund von Hintergrundwissen geeignet. (Prinzipien: “Müll rein – Müll raus” und “Es gibt keine bessere Daten als mehr Daten”)**


LERNENDE MASCHINEN 02.05.2017
INDUSTRIE 4.0
SPRACH-DIALOGE 09.05.2017
KÜNSTLICHE INTELLIGENZ
KI
BIG DATA
TEAM-ROBOTIK
AUTONOME SYSTEME 16.05.2017
ALTERS-ASSISTENZ
SMART SERVICE
SICHER-HEIT
EMOTION & VERHALTEN

Vielen Dank für Ihre Aufmerksamkeit