

Spezielle Themen der KI

NLP

Natural Language Processing

Marc Erich Latoschik
AI & VR Lab, Faculty of Technology
University of Bielefeld
marcl@techfak.uni-bielefeld.de

Was ist Sprache?

- Menschliche Sprache
 - universelles Kommunikations- und Speichermedium
 - Interaktion zw. Menschen über verschiedene „Kanäle“ von denen Sprache besonders dominant scheint
 - Wissensvermittlung, kollektiver Erkenntnisgewinn
- Sprache und Denken sind eng verknüpft
 - Kategorisierung, Benennung, Abstraktion
 - Erkenntnisgewinn bei Arbeit an sprachlicher Formulierung
 - mentale Abläufe menschlicher Sprachverarbeitung nicht bekannt

Natürliche Sprachen – formale Sprachen

- natürliche Sprache
 - menschliche Sprachen (Deutsch, Englisch...)
 - sehr mächtig in den Ausdrucksmöglichkeiten (kann „alles“ beschreiben)
 - oft mehrdeutig: Ambiguitäten
- formale Sprache
 - Computersprachen, Logiken (Aussagelogik, Prädikatenlogik,...)
 - einfach zu verstehen aber eingeschränkte Ausdrucksfähigkeit (fachspezifische Benutzung),
 - eindeutig definiert

Spezielle Themen der KI

NLP 1

Kognitionswissenschaftliche Betrachtungen: Sprachverstehen als Informationsverarbeitung

Marc Erich Latoschik
AI & VR Lab, Faculty of Technology
University of Bielefeld
marcl@techfak.uni-bielefeld.de

Sprachverstehen

Was ist Sprachverstehen?

- Prozess oder Funktion welcher physikalische Form einer Äußerung auf die Repräsentation ihrer Bedeutung abbildet.
- Kognitive Prozesse des Sprachverstehens werden (im Kognitionswissenschaftlichen Paradigma der Informationsverarbeitung) als Berechnungsvorgänge und Manipulation von Information verstanden.

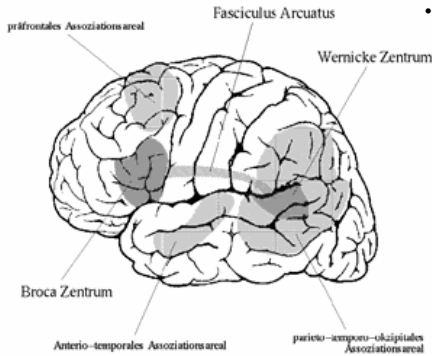
Sprachverstehen als Interaktion von Wissenssystemen

Zwei Typen von Kenntnis- und Wissenssystemen lassen sich im Kontext des Sprachverstehens unterscheiden:

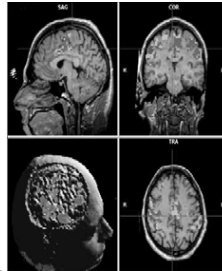
1. Allgemeine Kenntnissysteme mit kognitivem Status außerhalb der Sprachverarbeitung
 - Wissen über Entitäten
 - Wissen über Zusammenhänge in der Welt, etc.
2. Spezifische Kenntnissysteme im Rahmen der Sprachverarbeitung:
 - Phonologisches Wissen
 - Syntaktisches Wissen
 - Lexikalisches Wissen

Betroffene Disziplinen

Neurolinguistik - Sprache und Gehirn



- Lokalisierung der Hirnfunktionen
 - Messung aktiver Neuronenareale mithilfe von MRI (aktive Hirnareale bei Fingertapping-Aufgabe)
 - Untersuchung von Störungen/Aphasien z.B. bei Verletzungen im Gehirn
- Organisation der menschlichen Sprechfähigkeit
 - Hirnareale für die Sprachverarbeitung
 - Verstehen und Erzeugen von Sprache im Broca- und Wernicke-Zentrum



7

Betroffene Disziplinen

- Linguistik:
 - Sprachspezifische Kenntnissysteme.
 - Sprachübergreifende Theorie der strukturellen Regularitäten der Sprache.
- Psycholinguistik:
 - Experimentell fundierte Prinzipien der Verarbeitung sprachlichen Wissens die sprachliches Verhalten determinieren.
- Kognitionswissenschaft:
 - Annahmen über die Repräsentation benötigten Wissens und deren Auswirkungen auf die Verarbeitung unter Zugrundelegung dedizierter *parametrisierter* Verarbeitungsprinzipien für *empirisch überprüfbare* Hypothesen.

8

Betroffene Disziplinen

„ ... , as psycholinguistics has become more computational, and therefore more explicit, researchers with different perspectives are increasingly able to translate their metatheoretical disagreements into empirical questions.“
[Tanenhaus, 1988, S.30]

- Detaillierte Spezifikation eines Verarbeitungsmodells.
- KI und Theoretische Informatik entwickeln Werkzeuge für kognitive Modellierung hochkomplexer Sprachverarbeitungsprozesse.
- Korrektheit, Handhabbarkeit und Berechenbarkeit schränken gemeinsam mit den psychologischen, linguistischen und neurophysiologischen Rahmenbedingungen die Klasse der möglichen Modelle menschlicher Sprachverarbeitung ein.

9

Betroffene Disziplinen

Philosophie:

- Rolle oder Funktion der Sprache als Vermittler zwischen Mensch und Umwelt.
(Denken \leftrightarrow Sprache \leftrightarrow Wirklichkeit).
[Austin, 1962; Grice, 1975; Searle, 1969]
- Rolle der Sprache im menschlichen Handeln.
- Rolle der Sprache als wirklichkeitskonstituierende Strukturierungshilfe.
[Hörmann, 1976]

10

Betroffene Disziplinen

Künstliche Intelligenz

- Zentrale Frage der KI
 - Sprachverstehen: Extraktion von Wissen/Bedeutung aus der Sprache (geschrieben oder gesprochen)
 - Interaktion mit dem Computer
- Sprache und „menschliche“ Intelligenz
 - Intelligenzbegriff eng verknüpft mit Sprache
 - Vorschlag zum operationellen Nachweis von Intelligenz: Turing Test
- Verarbeitung natürlicher Sprache (natural language processing NLP)
 - Sprachverstehen
 - Spracherzeugung
 - Maschinelle Übersetzung
 - gesprochene oder geschriebene Sprache
 - Einzelne Worte, Sätze, lange Texte

11

Geschichte der Sprachverarbeitung

1791: Erste Sprechmaschine von Wolfgang von KEMPELEN

1956: Schaffung der Theorie der formalen Sprachen (außerhalb der KI) durch CHOMSKY.

60er

- NLP mit einfachen Mustertransformationen. Prototyp: ELIZA (Weizenbaum)
- Wissensrepräsentation mit Hilfe semantischer Netze (Quillian)

70er

- Effektive Grammatikformalismen (ATN-Grammatiken)
- SCHOLAR - Beantwortung von Fragen über Südamerika
- LUNAR (Wood) - Auskunftssystem über Gesteinsproben
- SHRDLU (Terry Winograd)
- dynamischer Mustervergleich, dynamische Programmierung

Beispiel für eine beschränkte Mikrowelt: Die „Klötzchenwelt“ von SHRDLU

80er

- statistische Modellierung (Hidden-Markov-Modelle)
- Diktiermaschine DRAGON DICTATE:
- SPICOS II - vollständiges Dialogsystem zur Datenbankabfrage

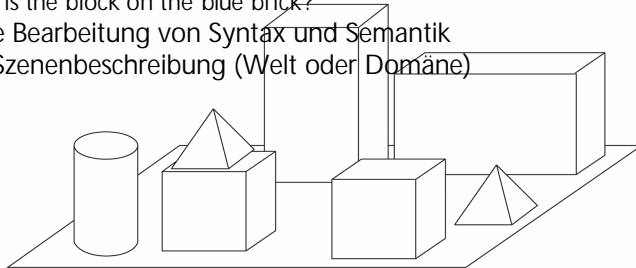
90er

- TANGORA - sprecherabhängiges automatisches Diktiersystem,
- EVAR - IC-Zugauskunft

12

SHRDLU – a blocks world [Winograd, 1972]

- SHRDLU: Konversation über eine Blockwelt.
 - What is sitting on the red block?
 - What shape is the blue block on the table?
 - Place a green pyramid on the red brick.
 - Is there a red block? Pick it up.
 - What color is the block on the blue brick?
- Modell für die Bearbeitung von Syntax und Semantik
- Vollständige Szenenbeschreibung (Welt oder Domäne)



13

Turing Test

I propose to consider the question "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." [Turing, 1950]

"Turing Test" soll untersuchen ob eine Maschine intelligent ist

Vorschlag von Turing: Imitationsspiel

- Mann A, Frau B, Fragesteller C
- C kann über Terminal mit A und B kommunizieren (räumlich getrennt) und Fragen stellen
- Ziel: C soll feststellen wer Mann und wer Frau ist (Kommunikationspartner mit X und Y bezeichnet), z.B "X ist A" und "Y ist B"

Turing Test für Intelligenz

- Fragesteller ist mit Person und Computer über Terminal verbunden
- Wenn das Programm den Befrager glauben machen kann, dass es ein Mensch ist "muß" es intelligent sein

Kritik am Turing Test: "Betrug" möglich, "Chinesisches Zimmer" (Searl)

14

Versuch eines Turing-Tests

Loebner Preis – Wettbewerb im Stil des Turing Test

- Kommunikation eines Fragestellers mit Menschen und Computerprogrammen über eine bestimmte Zeit
- Gewinner des Loebner-Preises 2003: Jabberwock
- Zukünftiges Preisgeld \$100.000 für eine audio-visuelle „Täuschung“



Einschränkungen

- Beschränkung auf bestimmte Gesprächsthemen
- Jury befindet über Zulässigkeit von Fragen
- Fragesteller bewertet wie „überzeugend“ er das Gespräch fand



Kritik: geschickte Täuschung ist effektiver als echtes Sprachverstehen

Konversationsprogramme

Programme, die mit Benutzern in natürlicher Sprache kommunizieren

- Chatbots
- Computer-Benutzer-Schnittstelle im Internet
- Assistenzsysteme, Kundenberatung

Eliza - Gesprächsprogramm nach dem Konzept von J. Weizenbaum

- Programm handelt/spricht wie ein Psychotherapeut
- Arbeitet mit Schlüsselwörtern und Kontextmustern
- Wissen ist abgelegt in Skripten
- **Kein Sprachverstehen**

Eliza-Effekt

- Eliza wirkt wie ein „verstehender“ Gesprächspartner
- Gesprächspartner interpretieren Bedeutungen in die Aussagen des Gegenüber
- hinein

ELIZA Programmierung

ELIZA SCRIPT FILE:

Sxxxxxxx - DEFINE SIGNON MESSAGE
 Txxxxx
 Txxxxx - DEFINE PAIR OF WORDS TO TRANPOSE
 Nxxxxx - RESPONSE FOR NULL ENTRY
 Mxxxxx - DEFINE RESPONSE FOR LATER USE OF "MY,
 Xxxxxxx - DEFINE RESPONSE FOR NO KEYWORD FOUND
 Kxxxxxx - DEFINE KEYWORD
 Rxxxxxx - DEFINE RESPONSE FOR PREVIOUS KEYWORD
 ...

SIGNON MESSAGE

SHI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT?
 SSO HOW ARE YOU DOING TODAY?
 ...

NULL RESPONSES

Nyou're gonna wear your enter key out if you keep that up.
 ...

TRANSPPOSITION TABLE

T ARE ,
 T AM ,
 T AM ,
 T ARE ,
 T WERE ,
 ...

NO KEYWORD FOUND

XPLEASE GO ON.
 XWHAT DOES THAT SUGGEST TO YOU?
 ...

DEFINE KEYWORDS AND RESPONSES

KCAN I ,
 RPERHAPS YOU DON'T WANT TO*.
 RDO YOU WANT TO BE ABLE TO*?
 KFRIEND ,
 RWHY DO YOU BRING UP THE TOPIC OF FRIENDS?
 RDO YOUR FRIENDS WORRY YOU?
 RDO YOUR FRIENDS PICK ON YOU?
 ...

17

Syntax, Weltwissen und Diskurs

Die Rolle der Grammatik

- Bis heute Kontroverse über die Rolle der Grammatik für NLP.
- Hypothese: Agrammatische Heuristiken steuern die syntaktische Verarbeitung [Bever, 1970; Matthews, 1979].
- Hypothese: Syntaktische Analyse nur nach fehlgeschlagener semantischer Analyse [Bever, 1970; Herrmann, 1990; Schank und Riesbeck, 1981]:

Syntax as last resort (SALR-Ansatz)

18

Die Rolle der Grammatik

- Garden-path-Phänomene als Argument des Fehlens eines grammatikbasierten Verarbeitungssystems:
(1) *The horse raced past the barn fell.*
- Sätze ähnlich (1) werden häufig von natürlichen Sprechern als ungrammatikalisch bezeichnet [Matthews, 1979].
- Aber: Warum finden Versuchspersonen vom syntaktischen Holzweg zurück? [Bresnan und Kaplan, 1982]

Die Rolle der Grammatik

- Reanalyse benötigt Wissen über alternative syntaktische Strukturen.
- Verarbeitungsmodelle mit „perzeptiven Strategien“ (etwa [Bever, 1970] oder [Fodor et al., 1974] können Reanalyse nicht erschöpfend erklären [Garnham, 1985].
- Strategie- oder bedeutungsgesteuerte Ansätze wurden oft mit Syntaxkomponente gekoppelt, um nach Fehlschlag der Analyse Syntax zu analysieren
 - SALR

Die Rolle der Grammatik

- SALR-Ansätze werden kaum noch vertreten.
- Psycholinguistische Experimente zeigen (etwa bei [D'Arcais, 1987]):
 - Erhöhter zeitlicher Verarbeitungsaufwand für in ihrer Bedeutung eindeutig ungrammatikalische Sätze...
 - ... auch wenn syntaktische Verletzung gar nicht bemerkt wird.
- Erklärung nur bei einer syntaktischen Analyse auch im einfachsten Fall

21

Die Rolle der Grammatik

(2a) *The red tomato ate the little boy.*

(2b) *The little boy ate the red tomato.*

- 2a und b werden für identisch gehalten (auch nach SALR-Ansatz)
- Wie kann neues ausgedrückt werden wenn Sprache ausschließlich aufgrund von Weltkenntnis verstanden würde?
- Verhältnis von Struktur und Bedeutung ermöglicht es, kreative neue Dinge über die Welt oder fiktive Welten auszudrücken!
- Was unterscheidet 2a und b von:
(2c) *Die rote Tomate aß der kleine Junge.*

22

Syntax, Weltwissen und Diskurs

Weltwissen und Diskurs

- Satzbedeutung ergibt sich nicht allein aus der Wortbedeutung
- Mentale Repräsentation ist nicht als Summe der Satzbedeutungen zu fassen.
- Nach [Crain und Steedman, 1985] werden strukturell ungewöhnliche Lesarten in Garden-path-Sätze leichter durch unterstützendes Weltwissen nahe gelegt:

(4a) *The teachers taught by the Berlitz method passed the test.*

(4b) *The children taught by the Berlitz method passed the test.*

- Wissen darüber, dass Kinder eher lernen als lehren erlaubt es in 4b „...taught by the Berlitz method...“ als reduzierten Relativsatz zu verstehen.