

INTERPRETATION VON GEBÄRDENSPRACHEINGABEN AM BEISPIEL EINER HOME-ENTERTAINMENT-STEUERUNG

Jens Lindemann

Brandenburgische Technische Universität Cottbus - Senftenberg
lindemann@tu-cottbus.de

Kurzfassung: Ausgehend von der Beschreibung des Bedeutungsbegriffs wird die semantische Analyse von lautsprachlichen Äußerungen in sprachverstehenden Systemen beschrieben. Die Umsetzung einer Sprachäußerung in eine Bedeutungsrepräsentation kann durch endliche gewichtete Transduktoren realisiert werden [1, 3, 11]. Die verwendeten Technologien lassen sich auch auf die Gebärdensprache übertragen. Dazu wird das Konzept des eigenen Forschungsvorhabens in dem Beitrag kurz vorgestellt. Am Beispiel einer Home-Entertainment-Steuerung soll die Semantikverarbeitung von laut- und gebärdensprachlichen Benutzeräußerungen vergleichend untersucht werden.

1 Semantische Analyse von Sprachzeichen

1.1 Die unmittelbare, wörtliche Bedeutung

Das Ziel der semantischen Analyse von natürlicher Sprache besteht in der Untersuchung der *wörtlichen*, vom Kontext der Sprechsituation unabhängigen, *inneren Bedeutung* eines syntaktischen Ausdrucks. Die maschinelle Extraktion von Bedeutung bzw. Wissen ist z. B.

- in Sprachdialogsystemen,
- zur Informationsgewinnung (speech information retrieval) und
- zur Sprachübersetzung in eine andere natürliche Zielsprache

notwendig. Die Berücksichtigung der konkreten Situation, in der eine sprachliche Äußerung durch einen bestimmten Sprecher artikuliert wird, ist hingegen Aufgabe der pragmatischen Analyse. Bei der semantischen Verarbeitung von Sprache soll die Beziehung zwischen einem vom Benutzer verwendeten Sprachzeichen und dem Zeicheninhalt bestimmt werden. Aus Sicht eines sprachverstehenden Systems wird zur Repräsentation der Bedeutung eines zu analysierenden Satzteils ein semantisches *Konzept* verwendet. Ein Konzept ist hierbei definiert als die kleinste Bedeutungseinheit. Für eine Sprachäußerung ergibt sich durch die bestehenden Relationen zwischen Konzepten eine Bedeutungsstruktur.

1.2 Der Bedeutungsbegriff nach Skinner

Der in [12] dargestellte behavioristische Ansatz zur Beschreibung der Bedeutung einer sprachlichen Äußerung basiert auf den von Burrhus Frederic Skinner veröffentlichten theoretischen Analysen verbalen Verhaltens. Die dabei aus Beobachtersicht durchgeführte Interpretation eines Kommunikationsprozesses liefert die *äußere Bedeutung* eines Sprechakts. Die Bedeutung von verbalen Verhalten (*Behavior B*) ist hierbei nicht in der kompositionellen Semantik seiner

einzelnen Konstituenten zu suchen, sondern in Analyse des bestehenden Kontextes (*Antecedents A* und *Consequences C*). Die äußere Bedeutung eines komplexen Sprachzeichens kann aus Beobachtersicht durch die festen, formal beschreibbaren Eigenschaften der Vorbedingungen *A* und den (beabsichtigten) Konsequenzen *C* angegeben werden. Wir bezeichnen dies als semantische Anker $\{a_i; c_j\}$ mit $a_i \in A$ und $c_j \in C$ [12]. Diese Definition des Bedeutungsbegriffs ermöglicht es, den tatsächlich vom Sprecher intendierten Inhalt einer Sprachäußerung in einer bestimmten Situation zu ermitteln.

1.3 Sprachverstehen in technischen Systemen – eine Übersicht

Um die wörtliche Bedeutung, die mit einem sprachlichen Ausdruck assoziiert ist, adäquat abzubilden, verwenden sprachverstehende, wissensverarbeitende Systeme semantische Modelle (siehe Abb. 1). Die Nutzung von syntaktischen Analyseinformationen aus ganzen Sätzen scheint dafür nicht der beste Lösungsansatz zu sein, denn die vom Benutzer verwendete Spontansprache weist doch oft

- eine andere Grammatik, im Vergleich zur geschriebenen oder vorgelesenen Sprache,
- Hässitationen, Wiederholungen, Selbstkorrekturen durch den Sprecher und
- unvollständige Sätze (Satzabbrüche)

auf. Die Ergebnisse der automatischen Spracherkennung können zudem fehlerbehaftet sein.

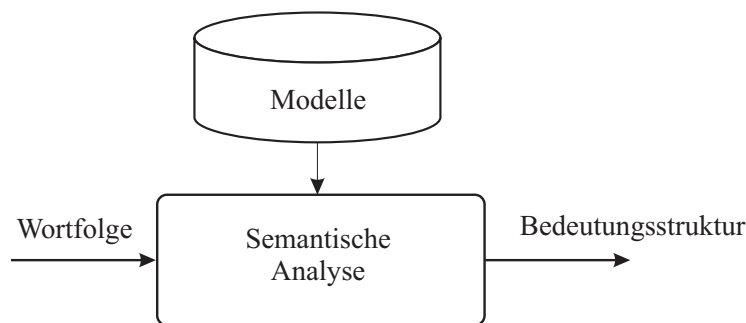


Abbildung 1 - Systemarchitektur der Entwicklungsumgebung

Heutige Sprachdialogsysteme beruhen meist auf regelbasierter semantischer Analyse [8] und betrachten einen eingeschränkten Gegenstandsbereich. Eine vorliegende Grammatik kann beispielsweise durch endliche Automaten beschrieben werden. Dabei können den Produktionsregeln anhand der Auftretenshäufigkeit in großen Datenbasen auch entsprechende Wahrscheinlichkeiten zugeordnet werden. Mit Hilfe dieser (stochastischen) Modelle kann das verbale Verhalten eines Nutzers in eine Bedeutungsrepräsentation (Zielsprache) überführt werden. Dabei besitzt die Notationssprache für die Bedeutungsbeschreibung wiederum eine eigene Syntax und Semantik. Die meisten in der Literatur beschriebenen Systeme verwenden Bedeutungsbeschreibungssprachen mit einer flachen Struktur. Zumeist erhält man als Ergebnis Sequenzen von Attribut-Wert-Paaren oder auch Semantische Frames, die mit den detektierten Merkmal-Wert-Paaren gefüllt wurden. Weiterhin können auch Baumstrukturen (vgl. [9]) zur Bedeutungsrepräsentation generiert werden. Aktuelle Systeme teilen die semantische Analyse in zwei nacheinander ausführbare Arbeitsschritte ein [1]:

1. Zuordnen der semantischen Konzeptmarkierungen zu der Wortfolge einer sprachlichen Äußerung mittels stochastischer Modelle
2. Merkmal-Wert-Extraktion aus der ermittelten Konzeptmarkierungssequenz

Diese Aufgaben können als Übersetzungsproblem betrachtet werden [1, 8]. Für sprachverstehende Systeme sind im Laufe der Zeit verschiedene Verfahren entwickelt worden. Das für die Entscheidung notwendige linguistische Wissen wird durch automatisches Lernen von stochastischen Übersetzungsmodellen erreicht. Eine wesentliche Grundlage für die Forschung und Entwicklungen wahrscheinlichkeitbasierter, sprachverstehender Systeme sind deshalb große Datenbanken. Je nach Umfang der Domäne können für das Training robuster Systeme mehrere tausend Sätze benötigt werden. Es bleibt anzumerken, dass für das Sprachverstehen auch die Ergebnisse einer pragmatischen Analyse zu berücksichtigen sind.

Übersetzung einer Sprachäußerung in eine Konzept(markierungs)folge

Für die Zuordnung der Werte zu den Merkmalen ist es zuerst notwendig, dass eine mögliche Wortfolge u_i mit $u_i = x_1, x_2, \dots, x_T$ in eine Konzeptfolge $v_i = y_1, y_2, \dots, y_L$ übersetzt wird. Eine Konzeptsequenz v_i ist eine endliche, geordnete Folge von Konzepten (semantischen Merkmalen) y_k mit der Länge $|v_i| = L$. Die Länge der Wörter einer Sprachäußerung $|u_i|$ muss nicht mit $|v_i|$ übereinstimmen. Die Konzepte y_k bestehen aus $N \geq 1$ Konzeptmarkierungen. Eine Konzeptmarkierungssequenz wird nachfolgend mit $\varphi_i = t_1, t_2, \dots, t_T$ bezeichnet. Ein Wort x_k aus der Sprachäußerung wird zu genau einer Konzeptmarkierung zugeordnet ($|u_i| = |\varphi_i|$). Die Reihenfolge der Konzeptmarkierungen eines Konzepts kann beispielsweise durch die BIO-Notation (Begin, Inside, Outside) gekennzeichnet sein [1]. Ausgehend von der in [12] verwendeten Beispieläußerung $u = \text{“Stell die Blumenvase auf die Kommode!“}$, ist in Bild 2 der prinzipielle Zusammenhang zwischen dem Quellsatz u , der Konzeptmarkierungssequenz φ und den daraus generierten Merkmal-Werte-Paaren dargestellt. Ein stochastischer Dekoder kann mit Hilfe der erstellten Modelle in der Arbeitsphase für eine gegebene sprachliche Äußerung u die wahrscheinlichsten Bedeutungsrepräsentationen bestimmen. Für diese Aufgabe können diskriminative und generative Modelle zum Einsatz kommen.

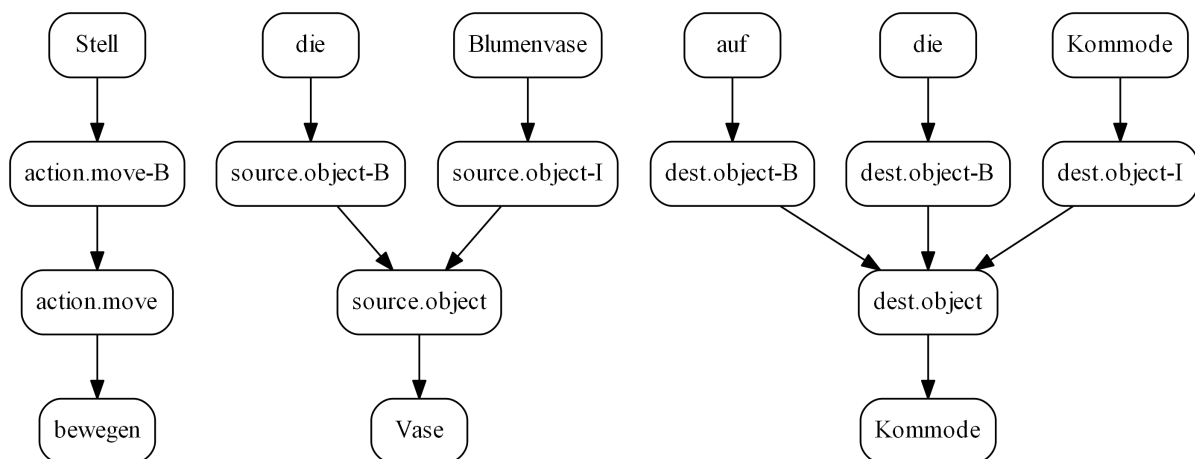


Abbildung 2 - Wortfolge u , Konzeptmarkierungssequenz φ , Merkmale und Werte (v. oben nach unten)

Diskriminative Verfahren bestimmen die Posteriorwahrscheinlichkeit $P(v|u)$ direkt. Generative Verfahren treffen ihre Entscheidung in Abhängigkeit einer a-priori gegebenen Bedeutungsstruktur. Formal ergibt sich dieser Zusammenhang durch das Bayes' Theorem mit

$$P(v|u) = \frac{P(u|v)P(v)}{P(u)}. \quad (1)$$

Die Entscheidungsfindung ist hierbei unabhängig von der Wahrscheinlichkeit einer gegebenen Wortsequenz $P(u)$. Nach [1] und [8] kommen für die semantische, stochastische Analyse in

sprachverstehenden Systemen vor allem

- generative (erzeugende) Modelle, wie
 - *Stochastic Parsing, Stochastic Finite State Transducers, Hidden Vector State Models, Dynamic Bayesian Networks, Standard Phrase-based Machine Translation*
- und diskriminative (unterscheidende) Modelle, wie
 - *Maximum Entropy Ansatz, Conditional Random Fields, Support Vector Machines*

zum Einsatz.

Merkmal-Wert-Extraktion

Möchte man die semantische Analyse mit einem konkreten Ergebnis abschließen, bietet sich im folgenden Schritt die regel- oder wahrscheinlichkeitsbasierte Bestimmung von Merkmal-Wert-Paaren an. Die Merkmale können dabei aus den Konzeptmarkierungen ermittelt werden. Für die Bestimmung der dazugehörigen Werte ist es notwendig, dass einem Konzept die dazugehörige Phrase der Quellsprache übergeben wird. Bei der Bestimmung der Werte müssen u. a. die unterschiedlichen sprachlichen Varianten, numerische Angaben (z. B. Datum und Telefonnummern) sowie verschiedene Zeitformen und Flexionen von Wörtern in eine zu definierende Stammform oder in eine allgemeinere Repräsentationsform gebracht werden. Für das automatische Training eines Normalisierungsmodells ist die Erfassung der normalisierten Werte bei der Annotation der Sprachdaten erforderlich. Alternativ kann die Merkmal-Wert-Extraktion auch durch die Anwendung deterministischer Regeln durchgeführt werden [1]. Die Normalisierung ist eine Grundvoraussetzung für die Konsistenz der Daten in sprachverstehenden Systemen. Nur so können in der obersten Systemebene Werte verglichen, mit Entitäten gerechnet und vorhandene Wissensmodelle angewandt und aktualisiert werden.

2 Realisierung der Übersetzung in eine semantische Struktur mittels endlicher gewichteter Transduktoren

Stochastische Modelle in Form von endlichen Automaten sind geeignet, um die Bedeutung einer Benutzeräußerung zu analysieren [1, 3, 11]. Der in aktuellen Forschungen vorgestellte Ansatz beruht auf der Vorstellung, dass die Bedeutung einer kompletten sprachlichen Äußerung durch die semantische Komposition seiner Konzeptelemente y_k beschrieben werden kann. Im Folgenden wird angenommen, dass als Ergebnis der Spracherkennung ein Worthypotheseograph vorliegt, der mehrere Erkennalternativen beinhaltet. Dieser gerichtete, azyklische Graph wird nachfolgend als Transduktor T_W bezeichnet. Ziel ist die Erstellung eines Transduktors T_{SA} zur semantischen Analyse. Dieser Transduktor muss ausgehend von T_W eine dazugehörigen Folge von Merkmal-Wert-Paare generieren. Eine Möglichkeit ist der in [1],[3] vorgestellte Ansatz. Darin wird ein T_{SA} durch Komposition (Hintereinanderausführung) von verschiedenen Transduktoren erstellt.

$$T_{SA} = T_W \circ T_G \circ T_{G2K} \circ T_{KSM} \circ T_{MWR} \quad (2)$$

Der *Generalisierungs-Transduktor* T_G übersetzt alle Wörter x_k , für die extern a-priori generalisierte Beschreibungen modelliert wurden, in Wortkategorien. Wenn für das Eingabewort x_k keine Verallgemeinerung verfügbar ist, gibt T_G das Eingabesymbol x_k wieder aus. Die Verwendung des Transduktors T_G hat u. a. den Vorteil, dass im Training nicht vorkommende, jedoch bedeutungstragende Wörter, wie z. B. numerische Angaben oder Eigennamen, fälschlicherweise einem *Füllwort-Modell* A_F zugeordnet werden. Die durch A_F akzeptierten Wörter werden

bei der semantischen Analyse in keine spezifische Bedeutungsmarkierung übersetzt. In φ werden diese Wörter durch ein Pseudokonzept, z. B. $t_k = null$, repräsentiert. Der Transduktor T_{G2K} übersetzt die Folge von Wörtern bzw. Wortkategorien in Sequenzen von möglichen Konzeptmarkierungen φ_i .

Bei der Verwendung von Transduktoren in einem Spracherkennungssystem erhält man auf Grundlage der akustischen Beobachtungsfolge \vec{o} hypothetische Sprachäußerungen u_i eines Nutzers. Um die beste Konzeptmarkierungssequenz $\hat{\varphi}$ zu einer akustischen Beobachtung \vec{o} zu bestimmen, müssen alle möglichen Wortfolgen und die dazugehörigen Konzeptmarkierungssequenzen betrachtet werden. Dieses Optimierungsproblem lässt sich durch

$$\hat{\varphi} \approx \arg \max_{\varphi_i} P(\vec{o} | u_i) P(u_i, \varphi_i) \quad (3)$$

ausdrücken [3]. Für die Auswahl der wahrscheinlichsten Abfolge von Konzeptmarkierungen t_k in dem Hypothesegraph wird ein stochastisches *Konzeptsprachmodell* (*KSM*) benötigt. Aus diesem wird der stochastische Transduktor T_{KSM} generiert, welcher das Verbundereignis (u, φ) berücksichtigt. Die Verbundwahrscheinlichkeit $P(u_i, \varphi_i)$ in der Formel 3 wird durch ein Folgenmodell ermittelt. Bei der Verwendung eines Trigramm-Modells, wie beispielsweise in [3], ergibt sich dieser Zusammenhang durch

$$P(u_i, \varphi_i) = \prod_{k=1}^N P(x_k, t_k | (x_{k-1}, t_{k-1}), (x_{k-2}, t_{k-2})). \quad (4)$$

Wenn das Wort x_k durch eine Wortkategorie repräsentiert werden kann, ist die Verwendung dieses verallgemeinerten Begriffs an Stelle von x_k bei der Berechnung in Gleichung 4 sinnvoll. Der Transduktor T_{MWR} übersetzt schließlich eine Konzeptmarkierungsfolge in eine Sequenz von Merkmal-Wert-Paaren (vgl. Kap. 1.3).

3 Semantische Analyse von Gebärdensprache

Um ältere Personen sowie Menschen mit Behinderungen nicht von der Verwendung von Dialogsystemen auszuschließen, sollte die Mensch-Maschine-Schnittstelle möglichst barrierefrei und multimodal gestaltet werden. Sprache ist die natürlichste Form der Mensch-Maschine-Interaktion. Ein intelligentes hierarchisches Dialogsystem sollte deshalb auch in der Lage sein Gebärdensprache zu perzipieren [6] und die Systemreaktion visualisiert darzustellen.

3.1 Gebärdensprache

Gehörlose Personen benutzen zur Verständigung untereinander die Gebärdensprache. In Deutschland wird die *Deutsche Gebärdensprache* (*DGS*) verwendet. Wie in der Lautsprache werden hierbei einzelne Gebärden durch grammatikalische Regeln zu Sätzen verknüpft. Somit können komplexe Sachverhalte ebenso gut und schnell wie in der gesprochenen Sprache ausgedrückt werden. Ein gebärdensprachliches Zeichen wird mit den Händen artikuliert und kann durch parallel ausgeführte bedeutungstragende nichtmanuelle Parameter, wie Gesichtsausdruck, Mundbild, Kopfhaltung, Blickrichtung und Ausrichtung des Oberkörpers ergänzt werden. Im Vergleich zur Lautsprache werden einzelne Phoneme einer Gebärde auch simultan artikuliert. Die Grammatik der DGS und der deutschen Lautsprache sind verschieden. Dies hat natürlich auch Auswirkungen auf die Modellbildung in sprachverstehenden Systemen. Im Vergleich zur Lautsprache wird in der DGS eine andere Wortstellung verwendet. Das Prädikat steht beispielsweise am Satzende und manuell ausgedrückte Adverbien, wie Zeit- und Ortsangaben, werden

am Satzanfang gebärdet. In der DGS wird kein Genus verwendet und es kommen auch keine Artikel vor. Weiterhin gibt es keine Hilfsverben, wie *sein*, *haben* und *werden*. In der DGS werden einfache Verben nicht flektiert, so dass auch kein trennbares Präfix vorkommt. Eine Konjugation von Verben wird jedoch durch einen Orts-, Personen- oder Objektbezug erzeugt. Bei abbildenden Verben, wie Subtitutorverben, wird die Hand zu Darstellung von imaginären Gegenständen verwendet, die in die Handlung einbezogen sind. Es können in der DGS auch Personen oder Gegenständen im Gebärdenraum „verortet“ werden. Durch die Indizierung ist es nachfolgend möglich, Beziehungen zwischen den Objekten zu beschreiben. Dies kann durch Raum- und Richtungsverben (z. B. geben, fragen) erfolgen, welche die konkrete Handlungsbeziehung zwischen der Quelle und dem Ziel angeben. Weiterführende Informationen zur Grammatik der DGS findet man u. a. in [2] und [10].

3.2 Home-Entertainment-System

Der für zukünftige Untersuchungen verwendete Gegenstandsbereich wurde bereits in [5] modelliert. Eine Home-Entertainment-Anwendung ermöglicht dem Benutzer folgende Funktionen mit audiovisuellen Medienobjekten (z. B. Film oder Fernsehsendung) auszuführen: *Aufnahme*, *Löschen*, *Wiedergabe* und *Kategorisierung der Medienbibliothek*. Dazu muss das System im Dialog mit dem Anwender die Aktion, das Medienobjekt und zusätzlich erforderliche Parameter eindeutig identifizieren können. In der Masterarbeit sind Systemstrukturen und -funktionen detailliert festgelegt worden. Für die stichprobenhafte Evaluation der erstellten statischen und dynamischen Modelle wurde ein Wizard-of-Oz-Experiment auf Basis von lautsprachlichen Benutzereingaben durchgeführt. An dem Versuch haben 12 Probanden (5x weiblich; 7x männlich) im Alter zwischen 17 und 45 Jahren teilgenommen. Die Testteilnehmer erhielten 11 verschiedene Aufgaben in Form einer elektronischen Programmzeitschrift. Darin wurde jeweils die auszuführende Problemstellung in geschriebener Sprache formuliert. Dies hatte zur Folge, dass die Wörter in der Aufgabenstellung vorzugsweise auch durch die Probanden verwendet wurden. Das zu identifizierende Medienobjekt wurde in der Programmzeitschrift hingegen meist mit typischen Fotos und Grafiken beschrieben und teilweise durch zusätzliche kurze Inhaltsbeschreibungen ergänzt.

3.3 Semantische Analyse von Gebärdensprache mittels endlicher Transduktoren

Gebärdensprachnotation

Bei der semantischen Interpretation soll einer erkannten Zeichenkette seine unmittelbare Bedeutung automatisch zugeordnet werden. Für die Verarbeitung von Gebärdensprache ist es notwendig, die visuellen Gebärdensprachzeichen durch ein geeignetes Notationssystem zu beschreiben. Bei der Gebärdenspracherkennung hat sich zur Verschriftlichung die *Glossennotation* etabliert. Allerdings existiert dafür leider keine einheitliche Kodierungsvorschrift. In der Hierarchie der Sprachzeichen sind Glossen etwa vergleichbar mit Wörtern der Lautsprache. Glossen sind mit lateinischen Großbuchstaben geschriebene Wörter, welche durch zusätzliche Zeichen ergänzt werden. Die Wörter einer Glosse stehen dabei meist für die Bedeutung eine Gebärde und werden in der Stammform notiert. Dies hat den Vorteil, dass die Komplexität des Modells für eine Wertnormalisierung geringer ausfällt. Für die Gebärdensprachverarbeitung mit endlichen Zeichenkettentransduktoren ist es erforderlich, dass die Glossennotation als Totalordnung beschrieben wird. Die Intension des in [12] verwendeten Beispielsatzes *“Stell die Blumenvase auf die Kommode!”* kann adäquat mittels Gebärdensprache ausgedrückt werden. Eine mögliche Notation zu einer gleichwertigen Gebärdensprachäußerung kann mit *KOMMODE BLUME+VASE (lh-cl-KOMMODE rh-cl-BLUME+VASE-AUF-KOMMODE)* angegeben werden. Das Wort *Blumenvase* wird durch die Komposition von den beiden Gebärden

BLUME und VASE ausgedrückt. Wenn mehrere Objekte in einem Satz auftreten, werden in der DGS große Objekte vor kleinen und unbewegliche Objekte vor beweglichen gebärdet. Nach der konkreten Benennung kann ein Gegenstand oder auch dessen Handhabung durch die Nutzung einer typischen Handform (Klassifikator) repräsentiert werden. In diesem Beispiel wird durch die letzte polymorphemische Gebärde die Bedeutung ausgedrückt, dass ein Objekt mit rundlichen Durchmesser auf der Kommode platziert wird. Der Endpunkt der Bewegung beschreibt die Zielposition. Die Visualisierung der Handhabung eines Objekts erfolgt hier durch die dominante Hand. In den eigenen Untersuchungen soll die manuelle Transkription ein Gebärdensprachexperte übernehmen.

Äußerungs-Bedeutungs-Transduktor

Für das zu entwickelnde System soll mit Hilfe eines Gebärdensprachexperten ein Transduktor erstellt werden, der für ausgewählte Funktionen erkannte Gebärdensprachäußerungen in *Merkmal-Werte-Relationen (MWR)* übersetzt. Zum Vergleich wird mit Hilfe des bereits vorliegenden Sprachmaterials aus [5] parallel eine Grammatik für lautsprachliche Benutzereingaben erzeugt. In [4] ist bereits exemplarisch die Konstruktion eines Transduktors auf Basis der in [5] gesammelten Sprachdaten für das Home-Entertainment-System beschrieben worden. Die in den geplanten Untersuchungen zu erstellenden Grammatiken werden aufgrund mangelnden Umfangs des Sprachmaterials und der Nichtverfügbarkeit eines geeigneten Gebärdenspracherkenners ungewichtet sein. Eine zu definierende Merkmalstruktur für die Medienobjekte kann aus dem in [5] erstellten Entity-Relationship-Modell abgeleitet werden.

Die Erfahrungen in [13] haben gezeigt, dass schon für kleinere Anwendungsbeispiele die Transduktoren sehr groß und unübersichtlich werden. Aus diesem Grund ist es sinnvoll zunächst die lokale Grammatik als Sammlung von Graphen zu schreiben. Hierbei bietet es sich an durch Generalisierungen *Platzhalter* in übergeordneten Graphen einzusetzen. Diese Platzhalter können im Gegensatz zu dem im Kap. 2 verwendeten Transduktoren T_G auch ganze Phrasen beschreiben. Da der Anteil der bedeutungstragenden Sprachzeichen in Gebärdensprachen vermutlich höher, und der Umfang des Wörterbuchs geringer ist, als für die deutsche Lautsprache, wird die Grammatik für die DGS kompakter ausfallen. Mit den in Kap. 2 beschriebenen endlichen Transduktoren können Spracheingaben nur über Umwege in eine komplexe MWR (Baumstruktur) überführt werden [6, 13]. Für die Generierung von Bedeutungsstrukturen und möglicherweise auch für die komplette semantische Verarbeitung (inklusive des Informationsstatus) in einem Kognitiven System scheinen Petrinetze ein geeignetes Mittel zu sein. Petrinetze sind formale Modelle für nebenläufiges Verhalten und können Nichtsequentialität eines menschlichen Gedankens besser modellieren. Mit der Definition von *Petrinetz-Transduktoren (PNTs)* und der notwendigen PNT-Operationen, wie z. B. Komposition, Summe und Produkt in [7] sind prinzipiell PNTs für die komplette Symbolverarbeitung in einem sprachverstehenden System verwendbar.

PNTs können als Verallgemeinerung der endlichen Transduktoren angesehen werden. Prinzipiell kann aus jedem endlichen Transduktor ein äquivalenter PNT erzeugt werden. Das Schalten einer Transition (Markierungswechsel) in einem Petrinetz ist dabei vergleichbar mit Zustandsübergang in einem endlichen Automaten. Die Transitionen sind mit einer Folge von Eingabesymbolen x des Alphabets X , bzw. mit Buchstaben y des Ausgabealphabets Y beschriftet. Durch nebenläufige Transitionen im Petrinetz erhält man nun eine *beschriftete partielle Ordnung*. Ein durch ein PNT erzeugter endlicher gerichteter Graph kann dabei als Bedeutungsrepräsentation (MWR) verwendet werden.

Aber auch für die Gebärdenspracherkennung, wo mehrere simultan artikulierte bedeutungstragende Parameter analysiert werden müssen, könnte die Verwendung von Petrinetz-Transduktoren die intuitivere Realisierungsmöglichkeit darstellen.

Literatur

- [1] DINARELLI, M.: *Spoken Language Understanding: from Spoken Utterances to Semantic Structures*. PhD thesis, 2010.
- [2] EICHMANN, H., M. HANSEN und J. HESSMANN: *Handbuch deutsche Gebärdensprache: Sprachwissenschaftliche und anwendungsbezogene Perspektiven*, Bd. 50 d. Reihe *Internationale Arbeiten zur Gebärdensprache und Kommunikation Gehörloser*. Signum, Seedorf, 2012.
- [3] HAHN, S., M. DINARELLI, C. RAYMOND, F. LEFEVRE, P. LEHNEN, R. D. MORI, A. MOSCHITTI, H. NEY, and G. RICCARDI: *Comparing stochastic approaches to spoken language understanding in multiple languages*. 19(6):1569–1583, 2011.
- [4] HUBER, M., C. KÖLBL, R. LORENZ und G. WIRSCHING: *Konstruktion von UMP-Transduktoren aus Wizard-of-Oz Daten*. In: WAGNER, P. (Hrsg.): *Elektronische Sprachsignalverarbeitung 2013*, Bd. 65 d. Reihe *Studientexte zur Sprachkommunikation*, S. S. 111 – 118. TUDpress, Dresden, 2013.
- [5] KARNAGEL, K.: *Semantische Modellierung von System-Benutzer-Interaktionen am Beispiel einer Home-Entertainment-Anwendung*. Masterarbeit, BTU Cottbus, 2012.
- [6] LINDEMANN, J.: *Semantische Verarbeitung von Gebärdensprache in intelligenten hierarchischen Sprachdialogsystemen*. In: HOFFMANN, R. (Hrsg.): *Elektronische Sprachsignalverarbeitung 2014*, Bd. 71 d. Reihe *Studientexte zur Sprachkommunikation*, S. 118–125. TUDpress Verlag der Wissenschaften Dresden, Dresden, 2014.
- [7] LORENZ, R., M. HUBER und G. WIRSCHING: *On Weighted Petri Net Transducers*. In: *Application and Theory of Petri Nets and Concurrency*, S. 233–252. Springer, 2014.
- [8] MACHEREY, K.: *Statistical methods in natural language understanding and spoken dialogue systems*. PhD thesis, 2009.
- [9] MILLER, S., R. SCHWARTZ, R. BOBROW, and R. INGRIA: *Statistical language processing using hidden understanding models*. In *Proceedings of the workshop on Human Language Technology*, pp. 278–282, 1994.
- [10] PAPASPYROU, C. E. A.: *Grammatik der deutschen Gebärdensprache aus der Sicht gehörloser Fachleute*. Signum, Seedorf, 2008.
- [11] RAYMOND, C., F. BÉCHET, R. D. MORI, and G. DAMNATI: *On the use of finite state transducers for semantic interpretation*. *Speech Communication*, 48(3–4):288–304, 2006.
- [12] WIRSCHING, G. and R. LORENZ: *Towards meaning-oriented language modeling*. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference*, pp. 369–374. 2013.
- [13] WIRSCHING, G. und M. WOLFF: *Semantische Dekodierung von Sprachsignalen am Beispiel einer Mikrofonfeldsteuerung*. In: HOFFMANN, R. (Hrsg.): *Elektronische Sprachsignalverarbeitung 2014*, Bd. 71 d. Reihe *Studientexte zur Sprachkommunikation*, S. 104–109. TUDpress Verlag der Wissenschaften Dresden, Dresden, 2014.