SEMANTISCHE VERARBEITUNG VON GEBÄRDENSPRACHE IN INTELLIGENTEN HIERARCHISCHEN SPRACHDIALOGSYSTEMEN

Jens Lindemann

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

Kurzfassung: In diesem Beitrag wird das Konzept des eigenen Forschungsvorhabens im Kontext der hierarchischen kognitiven Dialogsysteme vorgestellt. Die Benutzerschnittstellen solcher Systeme sollten multimodal gestaltet werden. In diesem Zusammenhang ist es für bestimmte Szenarien sinnvoll gehörlosen Personen die Spracheingabe mittels Gebärden zu ermöglichen, welche dann vom System analysiert und interpretiert werden müssen. Die Zuordnung einer sprachlichen Äußerung zu seiner Bedeutung wird im vorgestellten Ansatz sowohl für die Lautsprache als auch für die Gebärdensprache durch Transduktoren realisiert. Als Bedeutungsträger werden in unserem System *Merkmal-Werte-Relationen* verwendet. Die Informationsverarbeitung oberhalb der semantischen Ebene ist in dem Modell bedeutungsgetrieben, so dass sie aus derzeitiger Sicht für alle durch das System analysierbaren multimodalen Benutzereingaben identisch erfolgen kann.

1 Natürlichsprachliche Dialogsysteme

Die Sprache ist die bedeutendste und natürlichste Kommunikationsform. Auch technische Systeme sollten in der Lage sein im gegebenen Anwendungskontext sinnvolle freie sprachliche Benutzereingaben zu „verstehen“, so dass der Benutzer keine vorgegebenen Kommandophrasen lernen muss oder ihm System- oder Menüstrukturen bekannt sein müssen. Für die Akzeptanz eines solchen Systems als Kommunikationspartner sind neben einer hohen Spracherkennungs- und Sprachsynthesegüte auch ein vernünftiges Dialogmanagement sowie die Möglichkeit der Adaption des technischen Systems an den jeweiligen Nutzer nötig. Hierbei müssen nach [1] in zukünftigen Systemen verschiedene Gesichtspunkte berücksichtigt werden. Dies können zum Beispiel

- soziale Aspekte (z. B. Alter, Geschlecht, Herkunft (multilinguale Systeme)),
- sensorische und motorische Fähigkeiten des Benutzers,
- Verhalten und emotionaler Zustand des Anwenders,
- Umweltfaktoren, wie beispielsweise ein erhöhter Störgeräuschpegel,
- die kognitive Auslastung (Dialogdauer und -verlauf, Erfahrungen mit System),
- oder unerwartete Benutzereingaben sein.

Um ältere Menschen sowie Menschen mit Behinderungen nicht von der Verwendung von Dialogsystemen auszuschließen, sollte die Mensch-Maschine-Schnittstelle möglichst barrierefrei und somit auch multimodal (optisch, akustisch, haptisch) gestaltet werden. Bei Schwerhörigen und Gehörlosen besteht das Problem, dass akustische Informationen unzureichend oder nicht mehr wahrgenommen werden können. Bei gehörlosen Personen kann auch durch Einsatz von technischen bzw. medizinischen Hilfsmitteln keine ausreichende Sprachverstehbarkeit und somit auch –verständlichkeit erreicht werden. Dies führt gleichzeitig zu einer schlechten lautsprachlichen Artikulationsfähigkeit. Während Blinde und Sehbehinderte von lautsprachlichen Systemeingaben und akustischen Systemreaktionen profitieren, müssen Informationen für Gehörlose visualisiert dargestellt werden. Für diese Zielgruppe sollte deshalb auch die Möglichkeit der Gebärdenspracheingabe realisiert werden. Zudem ist der Einsatz von Gebärdensprachavataren für die Darstellung von komplexen Sachverhalten auf der Syntheseseite sinnvoll.

2 Gebärdensprache

Zur Kommunikation untereinander verwenden gehörlose Personen die Gebärdensprache. Wie bei den Lautsprachen gibt es auf der Welt unterschiedliche nationale Gebärdensprachen. In Deutschland z. B. wird die *Deutsche Gebärdensprache* (*DGS*) verwendet. Weiterhin gibt es innerhalb eines Landes auch regional unterschiedliche Sprachvarianten und Dialekte sowie Abweichungen zwischen verschiedenen Generationen von Gehörlosen. Zur Darstellung der Gebärdensprache werden visuell wahrnehmbare Sprachelemente verwendet. Gebärdensprachliche Zeichen werden mit den Händen artikuliert und durch parallel ausgeführte bedeutungstragende nichtmanuelle Parameter, wie Gesichtsausdruck, Mundbild, Kopfhaltung, Blickrichtung und Ausrichtung des Oberkörpers ergänzt. Die simultan ausgeführten nichtmanuellen Parameter dienen u. a. zur Präzisierung einer Gebärde, zur Bedeutungsunterscheidung, zur Kennzeichnung von Satztypen sowie zur Artikulation von nicht durch Gebärden darstellbaren Adjektiven und Adverbien [2], [3], [4]. Die simultan beobachtbaren manuellen und nichtmanuellen Ausdrucksmittel sind vergleichbar mit den Phonemen der Lautsprache.

Neben der Verwendung von konventionellen Gebärden können auch spontan neue, bildhafte Gebärden erzeugt werden. Solche produktiven Gebärden werden zum Beispiel zur Beschreibung von Formen, Bewegungsabläufen und räumlichen Verhältnisse verwendet. Unbekannte Begriffe, wie Fachtermini und Eigennamen können mittels des Fingeralphabets „buchstabiert“ werden. Weiterhin können Einzelgebärden auch vereinigt werden (*Inkorporation*). Als Beispiel kann der Ausdruck „VIER-STUNDEN“ angegeben werden. Hierbei erfolgt die Integration der Zahlhandform in die Gebärde STUNDE. Gebärdensprachen zeichnen sich durch ein hohes Maß an Simultanität aus. Komplexe Sachverhalte können mittels der Gebärdensprache ebenso gut wie in der gesprochenen Sprache vermittelt werden. Diese Form der Kommunikation ist zudem in ihrer Ausdrucksfähigkeit und Sprechdauer durchaus vergleichbar mit der Lautsprache [5].

Notationssysteme

Für die Sprachverarbeitung in technischen Systemen ist es erforderlich die gesprochenen Sprachzeichen in einer äquivalenten schriftlichen Form abzubilden. Für die Gebärdensprache existiert aktuell kein einheitliches Notationssystem. Basierend auf dem *Parametermodell* [6] gibt es linguistische Notationsformen, wie die *Stokoe-Notation* und das *Hamburger Notationssystem* (*HamNoSys*), mit denen die Gebärden phonetisch transkribiert werden können. Die im HamNoSys verwendeten Symbole, welche als Unicode verfügbar sind, repräsentieren die manuellen und nichtmanuellen Ausdrucksmittel der Gebärdensprache. Aufgrund der Vielzahl möglicher Gebärdensprachparameter wird der Schriftsatz stetig weiterentwickelt. Gerade die nichtmanuellen Komponenten sind in der aktuellen Version des HamNoSys noch unzureichend berücksichtigt worden. Ein häufig verwendetes Schriftsystem stellt die Glossen-Notation dar. Glossen sind mit lateinischen Großbuchstaben geschriebene Wörter, welche durch zusätzliche Zeichen ergänzt werden. Eine Glosse steht dabei meist für die Bedeutung eine Gebärde. Im Vergleich zu phonetisch motivierter Notation stellt dies also eine höhere Abstraktionsebene dar. Im Allgemeinen ist diese Notationsform auch für Nichtgebärdensprachler gut verständlich und leicht zu erlernen. Die zusätzlichen Symbole werden für grammatikalische Angaben, wie z. B. für Angaben zu Positionen im Gebärdenraum, Betonung, Negation oder Pluralbildung verwendet. Für die Glossen-Notation existiert leider keine einheitliche Codierungsvorschrift. Neben linguistischen Verschriftlichungsformen gibt es auch animationsorientierte Notationssysteme, wie z. B. *Signing Gesture Markup Language* (*SIGML*) sowie *Embodied Agent Behavior Realizer Script* (*EMBRScript*), welche zur Ansteuerung von Gebärdensprachavataren entwickelt wurden und zumeist auf XML basieren.

Syntax

Die Gebärdensprachzeichen werden unter Berücksichtigung von grammatikalischen Regeln zu Sätzen verknüpft. Der Satzbau von DGS unterscheidet sich von der Syntax der deutschen Lautsprache. Die Wortfolge in der Gebärdensprache ist relativ fest und spielt für die Bedeutung einer Aussage eine vergleichsweise untergeordnete Rolle. Grammatikalische Funktionen werden vorwiegend durch eine Abwandlung oder Hinzufügen einzelner Sprachparameter erreicht, nicht aber durch eine Änderung der Wortfolge. Ein wesentlicher Unterschied zur Lautsprache ist, dass das Prädikat am Satzende steht. Die beiden Satzglieder Subjekt und Objekt können jedoch auch in das Prädikat (durch Bewegungsrichtung, Ausführungsstelle oder Handform) inkorporiert werden [4]. Bei der Benennung mehrerer Objekte in einem Satz wird allgemein die Regel verwendet, dass große Objekte vor kleinen und unbewegliche Objekte vor beweglichen genannt werden. Zudem ist es üblich, dass wichtige Aspekte vor unwichtigen artikuliert werden. Manuell ausgedrückte Adverbien, wie Zeit- und Ortsangaben (z. B. „HIER“, „MORGEN“, „FRÜHER“), werden am Satzanfang gebärdet, wobei Zeit- vor Ortsbezeichnungen erfolgen. Abfolgen werden chronologisch dargestellt, wie in dem Beispiel 1 zu erkennen ist. Ein attributives Adjektiv folgt nach Substantiv oder wird simultan durch Mimik ausgedrückt [3].

Beispiel 1: Satz in Schriftsprache und Glossennotation für mögliche Gebärdensprachrealisierung

Ich gehe heute zur Universität, nachdem ich gefrühstückt habe.

HEUTE INDEX-ich FRÜHSTÜCKEN DANN UNIVERSITÄT ich-BESUCHEN-mi

In der DGS werden einfache Verben nicht flektiert. Eine Konjugation wird meist nur durch einen Orts- oder Personenbezug erzeugt. Bei dem Transferverb bzw. Richtungsverb „besuchen“ (siehe Beispiel 1) wird die Bewegung vom Subjekt zum Objekt ausgeführt [4].

3 Intelligente hierarchische Dialogsysteme

Für das an der TU Dresden entwickelte hierarchische System UASR (*Unified Approach for Speech Synthesis and Recognition*) zur gemeinsamen Sprachsynthese und –erkennung sind Weiterentwicklungen nach dem Vorbild der humanen Informationsverarbeitung geplant [7]. Ein Konzept für die Struktur eines intelligenten dynamischen hierarchischen Dialogsystems auf Basis von [7] ist durch Abb. 1 gegeben. Ein Ansatzpunkt ist die Implementierung einer bidirektionalen Informationsverarbeitung zwischen den Hierarchieebenen, wie sie in biologischen und sensomotorischen Modellen zu finden ist. Dabei können durch Spiegelung der Signalverarbeitung des Kommunikationspartners im eigenen Systemmodell Informationen zur Prädiktion auf der Analyseseite aus höheren Verarbeitungsebenen genutzt werden. Solch ein System besitzt also die Möglichkeit Objektverhalten „vorherzusagen“. Zudem kann auch die Synthese durch Feedbackinformation aus der Bottom-Up Richtung profitieren. Derzeitige Forschungsaufgaben beschäftigen sich auch mit der semantischen und pragmatischen Verarbeitung im UASR-System. Im Folgenden werden einige Anforderungen beschrieben, die aus unserer Sicht technische Systeme kognitiv machen. In der Literatur werden bereits Lösungsansätze hierfür aufgezeigt. Trotzdem gibt es in einigen Teilgebieten noch großes Entwicklungspotential. Ein kognitives System muss ein zielgerichtetes Verhalten aufweisen. Die optimale Systemreaktion ist ggf. mittels einer Bewertungsfunktion anhand der zu erwartenden Konsequenzen im Dialogverlauf auszuwählen. Zudem muss die Verhaltenssteuerung für eine durchgeführte Aktion die Erwartungswahrscheinlichkeiten für die nächste Benutzereingabe berechnen können. Neben der Möglichkeit Dialogstrategien anzupassen, besitzen intelligente Systeme die Fähigkeit zum Schließen und Folgern. Somit können beispielsweise mehrdeutige Eingaben unter Verwendung von Kontextinformationen aufgelöst bzw. interpretiert werden. Weiterhin müssen sich solche technischen Systeme an die Umgebung und an das Kommunikationsobjekt anpassen können. Durch die Fähigkeit des Systems zu Lernen, können Wissensmodelle adaptiert und somit eine höhere Robustheit

erreicht werden. Ein kognitives Dialogsystem muss nicht nur Entscheidungen unter Unsicherheit treffen, sondern auch seine Umgebung ausreichend wahrnehmen können. Dafür ist die Implementierung eines multimodalen Systems nötig, welches akustische, visuelle und haptische Information vom Benutzer aufnehmen kann. Sinnvoll ist eine parallele Erkennung und Interpretation von unterschiedlichen Eingabeformen durch das technische System. Dies führt im Allgemeinen zu einer erhöhten Interaktionssicherheit, Natürlichkeit und Flexibilität.

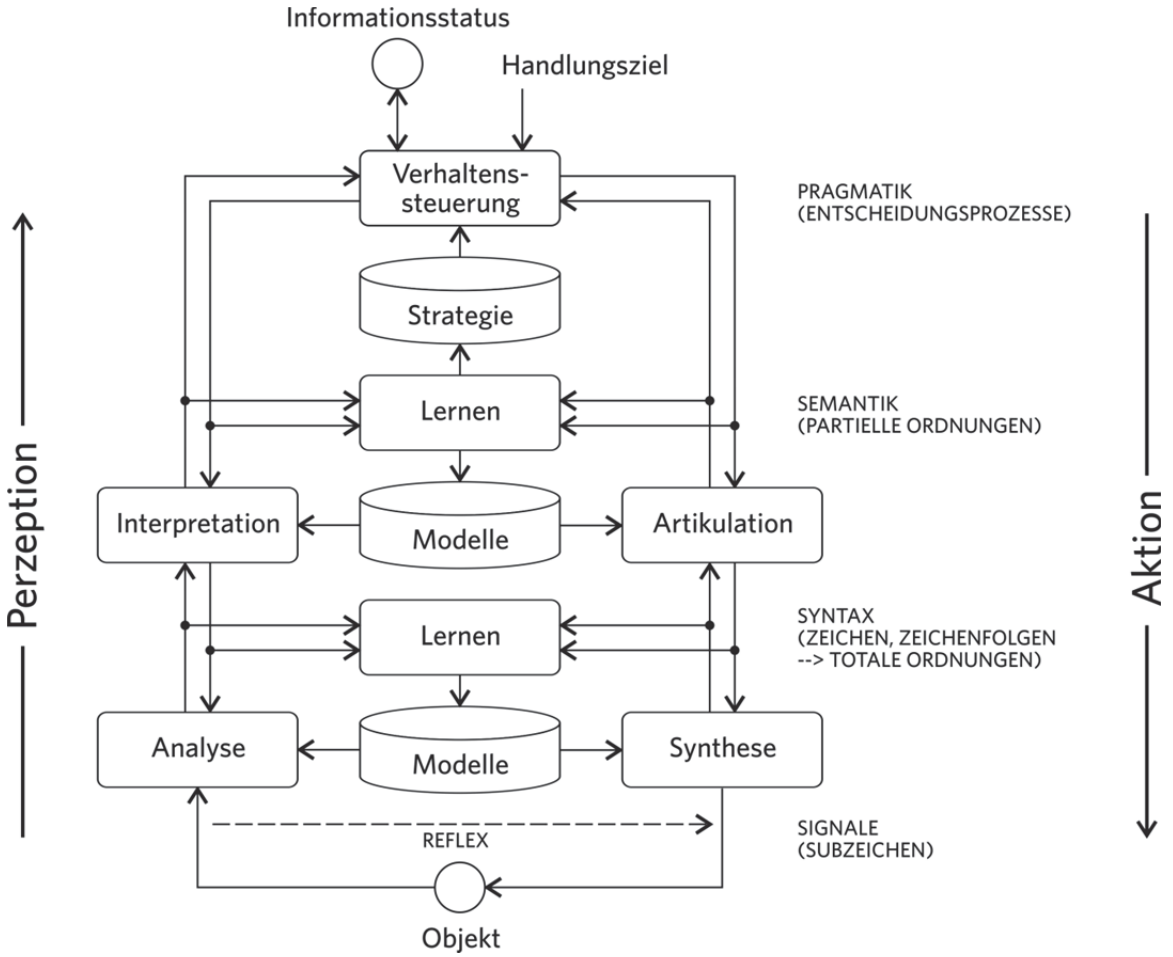

Abbildung 1 – Allgemeine Systemarchitektur eines kognitiven hierarchischen Dialogsystems

In den verschiedenen Verarbeitungsebenen kann durch die aufgebauten Parallelwelten somit möglicherweise eine Entscheidung unter geringerer Unsicherheit getroffen werden. Ein intelligentes Dialogsystem sollte auch gebärdensprachliche Eingaben analysieren und „verstehen“ können und adäquat darauf reagieren. Natürlich sind für die Analyse und Synthese von Gebärdensprache andere Modelle als für die Lautsprache notwendig, da sich beide Kommunikationsformen grundsätzlich unterscheiden. Aktuelle Systeme zur sprecherunabhängigen kontinuierlichen Gebärdenspracherkennung arbeiten ebenfalls nach dem Prinzip der Lautsprachanalyse [8], [9]. Dazu sind eine Klassifikation von geeigneten visuellen Merkmalvektorfolgen für die Erkennung von manuellen und nicht-manuellen Gebärdensprachparametern und die statistische Übersetzung in eine Wortfolge (Text) erforderlich. Problematisch für die Übersetzung ist die Notwendigkeit von entsprechenden großen Mengen an annotierten bilingualen Sprachdatenbasen [10]. Auf der Syntheseseite wird die ausgewählte Systemantwort idealerweise für Gehörlose durch einen Avatar (virtuellen Menschen) oder gegebenenfalls auch mittels Piktogrammen dargestellt. Prinzipiell kann auch Text direkt auf einem Display präsentiert werden. Allerdings ist das Leseniveau bei dieser Zielgruppe oft nur mäßig ausgeprägt, so dass bei umfangreichen Formulierungen eventuell Akzeptanzprobleme auftreten.

4 Gebärdensprachinterpretation in intelligenten technischen Systemen

In den geplanten eigenen Untersuchungen soll für ein konkretes Anwendungsszenario die semantische Verarbeitung von Gebärdensprache realisiert werden. Parallel dazu ist für dieselbe eingeschränkte Domäne die Analyse von lautsprachlichen Benutzereingaben angedacht. Die Funktionalitäten des Anwendungssystems sowie die Erstellung einer Wissensbasis (Weltmodell) sind unabhängig von der Eingabesprache und können deshalb für Laut- und Gebärdensprache gemeinsam definiert und modelliert werden. Zur Verarbeitung und Modellierung von Daten, wie HMMs (Hidden Markov Models), Grammatiken und Lexika, werden im UASR-System einheitlich *Finite State Transducers* (*FST*) verwendet. Auf Basis dieser Systemtechnologie und der Möglichkeit der Verwendung der Automatenalgebra kann möglicherweise auch die Interpretation der Nutzeräußerung erfolgen. Ein Spracherkenner liefert als Ergebnis der Analyse einer artikulierten sprachlichen Äußerung (*Utterance*) des Benutzers einen gewichteten Text- bzw. Wortphrasenhypothesegraph, jedoch keine Angaben zur möglichen Bedeutung (*Meaning*) der Spracheingabe. Bei der semantischen Interpretation soll nun der erkannten Zeichenkette seine unmittelbare Bedeutung durch sogenannte *Utterance-Meaning Pairs (UMP)* zugeordnet werden. Die Bedeutung einer Aussage ist unserer Meinung nach nicht in der kompositionellen Semantik der Bedeutungen seiner einzelnen Konstituenten zu suchen, sondern in der beabsichtigten Wirkung (*Consequences*) der gesamten sprachlichen Äußerung, unter Berücksichtigung von Kontextinformationen (*Antecedents*), wie z. B. des aktuellen System- bzw. Dialogzustands. Die Intention des Benutzers, welche er durch sein Verhalten bei konkret definierten Vorbedingungen zum Ausdruck bringt, könnte auch für die Adaption der semantischen Modelle verwendet werden. Der beschriebene Ansatz basiert auf der Theorie von Burrhus Frederic Skinner zur Analyse sprachlichen Verhaltens (*Behavior*) (vgl. [11]). Es gibt natürlich eine Vielzahl von konkreten Äußerungen für das Erreichen eines Kommunikationsziels. In der Domäne des Dialogsystems hat also unterschiedliches Benutzerverhalten, bei identischen Antecedents (Vorbedingungen) und denselben Konsequenzen, für das System dieselbe Bedeutung. Auf diesem Weg wäre es möglich, auch neue Äußerungen mit derselben intentionalen Bedeutung zu erlernen.

Die semantische und pragmatische Repräsentation der sprachlichen Äußerungen wird in unserem hierarchischen Systemmodell durch *gewichtete Merkmal-Werte-Relationen (gMWR)* [12] realisiert. Dies sind partielle Ordnungen über eine Menge von Merkmalen und Werten. Die erkannten Bedeutungsträger werden durch Unifikation in den bestehenden Informationsstatus des Systems (vgl. Abb. 1) eingetragen. Dieser Informationsstatus beinhaltet eine Menge von Merkmal-Werte-Paaren und liefert zudem die Interaktionshistorie. Aus diesem Informationsstand kann die Verhaltenssteuerung entnehmen, ob durch Benutzereingaben eine Systemfunktion mit der zu ihrer Ausführung benötigten Parametern eindeutig identifizierbar ist. Mit Hilfe des Informationsstatus kann das System also zielgerichtet auf seine Umgebung einwirken (Aktion). Entweder ist es möglich eine angeforderte Funktion auszuführen, oder die dafür fehlenden Informationen müssen vom Benutzer durch Rückfragen eingeholt werden.

Oberhalb der semantischen Ebene werden in unserem Systemmodell für Laut- und Gebärdensprache dieselben nichtsequentiellen Bedeutungsträger verwendet. Dies führt zu der These, dass die Informationsverarbeitung in der semantischen und pragmatischen Ebene des Systems unabhängig von der perzipierten Sprach- bzw. Eingabeform durchgeführt werden kann. Da derzeit kein Gebärdenspracherkenner verfügbar ist, soll für die semantische Analyse zunächst eine manuelle Transkription gebärdensprachlicher Äußerungen der Probanden durch einen Gebärdensprachdolmetscher oder –experten erfolgen. Dieser stellt jedoch im Normalfall einen idealen Spracherkenner dar, welcher im Allgemeinen eine eindeutige und ungewichtete Entscheidung liefert. In Abhängigkeit vom Notationssystem verwendet man für die Verschriftlichung ein spezielles Alphabet X. Als Notation der erkannten Gebärdensprach-

äußerung $u \in X^*$ bietet sich eine Glossenumschrift (siehe Kap. 2) an. In der Hierarchie der Sprachzeichen sind Glossen etwa vergleichbar mit den Wörtern in der Lautsprache.

Für das zu realisierende System ist in unserem Fall eine lokale Grammatik zu entwickeln, welche möglichst alle sinnvollen Sätze in der Domäne akzeptiert und unter Berücksichtigung einer ungeordneten Menge von Antecedents deren Bedeutung ausgibt. Diese Interpretation der Textnotation von gebärdensprachlichen Äußerungen soll in der Systemapplikation durch gewichtete Transduktoren [13], [14], [15] erfolgen. Ein Transduktor bildet im Allgemeinen eine endliche Zeichenkette $u \in X^*$ der Sprache $L_X \subseteq X^*$ in eine Zeichenfolge $v \in Y^*$ der Zielsprache $L_Y \subseteq Y^*$ ab. Die Zeichenfolgen bestehen aus Eingabesymbolen x des Alphabets X bzw. aus Buchstaben y des Ausgabealphabets Y, wobei das leere Zeichen ε jeweils Bestandteil der Alphabete X und Y ist. Schon für kleinere Anwendungsbeispiele können die UMP-Transduktoren sehr groß und unübersichtlich werden. Deshalb 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 müssen dann durch ihre dazugehörigen Transduktoren spezifiziert werden. Die dadurch entstehenden Subgraphen können dann modular als Bestandteil anderer Graphen eingesetzt werden. Die auf diesem Weg definierten mikrolokalen Grammatiken können letztendlich wieder zu einem gesamten UMP-Transduktor vereinigt werden. Zudem muss bei der Generierung des UMP-Transduktors nur der Gegenstandsbereich modelliert werden, der in dieser beschränkten Welt sinnvoll erscheint. Die Erstellung der lokalen Grammatik soll mit Hilfe von *Wizard-of-Oz* Experimenten manuell realisiert werden [16]. Aus diesem Grund sollte für prinzipielle Untersuchung des hier vorgeschlagenen Ansatzes die Domäne des Dialogsystems zunächst stark eingeschränkt werden. In den Wizard-of-Oz Experimenten wird ein funktionierendes System simuliert und Probanden müssen hierbei unter Angabe bestimmter *Antecedents* verschiedene vorgegebene Zielvorstellungen (*Consequences*) erreichen. Dies dient zum einen zum Sammeln verschiedener möglicher sprachlicher Äußerungen (*Behavior*) und zum anderen auch zur Überprüfung und ggf. Verfeinerung der Modellierung des zu erstellenden Anwendungssystems.

Ein UMP-Transduktor hat bei der Perzeption die Aufgabe aus einer sequentiellen textlichen Repräsentation der Benutzereingabe alle möglichen intentionalen Bedeutungen als partielle Worte zu generieren. Falls sich nicht aus der Gewichtung eine finale Bedeutung des Systems ergibt, muss diese gezielt beim Benutzer erfragt werden. Weiterhin können UMP-Transduktoren auf der Perzeptionsseite auch für die zu artikulierenden semantischen Daten alle passenden Wortfolgen ermitteln. Diese Zeichenketten können dann für die nachfolgende Synthese verwendet werden.

Der aktuelle Stand des Konzepts zur Umsetzung einer semantischen Interpretation für das zu realisierende System ist in [17] beschrieben. Bisher werden durch FSTs Zeichenketten erzeugt, welche die Informationen für die Erstellung gewichteter MWRen zu den analysierten Sprachäußerungen beinhalten. Die erkannten semantisch bedeutenden Bestandteile werden hierbei zunächst in der Reihenfolge von links nach rechts aufgeführt, in der sie erkannt wurden. Zu beachten ist, dass bei partiellen Worten, die als Interpretationsergebnis der UMP-Transduktoren erzeugt werden, die Reihenfolgeinformation nicht mehr erhalten bleibt. Dies ist auch in den meisten Fällen sinnvoll. Betrachtet man den Satz in Schriftsprache aus dem Beispiel 2, so hat dieser unter Annahme der gleichen Vorbedingungen dieselbe Bedeutung, wie der Satz „*Von Berlin nach Dresden möchte ich morgen fahren.*“, obwohl sich die Reihenfolge der Satzglieder unterscheidet. In einigen Fällen benötigt man jedoch die Information, in welcher Reihenfolge bestimmte sprachliche Zeichen genannt wurden. In dem aufgeführten Beispiel 2 ist dies erforderlich, um für die Routenberechnung den entsprechenden Bezug zu den Ortsangaben zu erhalten. Das Transferverb FAHREN definiert in diesem Beispiel durch die Verortung im Gebärdenraum den Start- und Zielpunkt. Man kann sich hierbei auch die komplexere Routenanfrage vom Startpunkt A über B und C zum

Ziel D vorstellen. Dabei kann es ein großer Unterschied sein, ob man erst über B und dann über C oder erst über C und dann über B fährt. Aus diesem Grund wurden in der Ausgabezeichenkette des Transduktors alle gleichnamig beschrifteten Merkmale, welche den gleichen Vorgängerknoten besitzen, für die Weiterverarbeitung hintereinander aufgeführt [17].

Beispiel 2: Glossennotation und Zeichenkette mit semantischer Information für sprachliche Äußerung

geschriebene Sprache s:

Ich möchte morgen von Berlin nach Dresden fahren.

Glossennotation u für eine mögliche Realisierung von s in Gebärdensprache:

MORGEN INDEX-ich WUNSCH BERLIN INDEX-li DRESDEN INDEX-re li-FAHREN-re

möglicher ungewichteter FVR-String v für Gebärdensprachäußerung:

[*date_rel*[*tomorrow*]][*loc*[*Berlin*][*loc*[*Dresden*]]]*route_abs*[*direction_abs*[*1to2*]]

Für eine äquivalente und übersichtliche Darstellung des FVR-Strings kann man mit Hilfe der Klammerzeichen eine Baumstruktur erzeugen. Für die Merkmal-Werte-Relation im Beispiel 2 erhält man den in Abbildung 2 dargestellten beschrifteten gerichteten Graphen. Leider kann ein FST diese partiellen Worte nicht direkt erzeugen, so dass für die technische Realisierung wahrscheinlich *Petrinetz-Transduktoren* (*PNTs*) zu verwenden sind. Diese können als Verallgemeinerung der FSTs angesehen werden. Entsprechende algebraische Operationen, wie die Komposition, wurden z. B. bereits in [15] definiert.

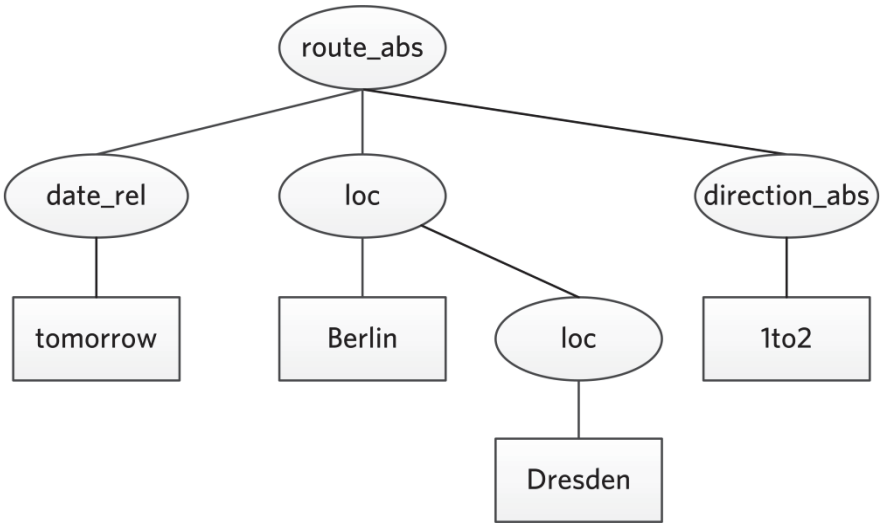

Abbildung 2: Baumstruktur zur FVR-Zeichenkette aus Beispiel 2

Zu berücksichtigen ist hierbei, dass diese Form der semantischen Präsentation die unmittelbare Bedeutung der Äußerung u angibt. Aus diesem Grund erfolgt im Beispiel 2 auch noch keine absolute Datumsangabe. Die notwendige Substitution in der MWR wird im dynamischen Dialogsystem erst in einem höheren Verarbeitungsschritt mit Hilfe von Weltwissen realisiert.

5 Ausblick

In diesem Beitrag wurde ein Konzept für die semantische Verarbeitung von sprachlichen Äußerungen vorgestellt. Darauf aufbauend soll in der nächsten Zeit ein System zur gemeinsamen Interpretation von Laut- und Gebärdensprachäußerungen erstellt und praktisch erprobt werden. Dazu sind neben der Entwicklung von effizienten Algorithmen noch weitere Untersuchungen zum unüberwachten Lernen, zur Gewichtung von semantischen Trägern und zur Informationsverarbeitung auf der pragmatischen Ebene notwendig.

Literatur

[1] Schmitt, Alexander; Minker, Wolfgang: *Towards adaptive spoken dialog systems.* New York, NY: Springer, 2013

[2] Boyes-Braem, Penny: *Einführung in die Gebärdensprache und ihre Erforschung.* Hamburg: Signum, 1990.

[3] Happ, Daniela; Vorköper, Marc-Oliver: *Deutsche Gebärdensprache. Ein Lehr- und Arbeitsbuch.* Frankfurt am Main: Fachhochsch.-Verl, 2006.

[4] Papaspyrou, Chrissostomos et al: *Grammatik der deutschen Gebärdensprache aus der Sicht gehörloser Fachleute.* Seedorf: Signum, 2008.

[5] Lindemann, Jens: *Vergleich der Produktionsgeschwindigkeit von Laut- und Gebärdensprache.* In: Matthias Wolff (Hg.): Elektronische Sprachsignalverarbeitung 2012. S. 236–243: TUDpress. Dresden (Studientexte zur Sprachkommunikation, 64), 2012.

[6] Dotter, Franz; Holzinger, Daniel: *Typologie und Gebärdensprache: Sequentialität und Simultanität.* In: Sprachtypologie und Universalienforschung (48), S. 311–349, 1995.

[7] Wolff, M.; Römer, R.; Hoffmann, R.: *Hierarchische kognitive dynamische Systeme zur Sprach- und Signalverarbeitung.* In: Matthias Wolff (Hg.): Elektronische Sprachsignalverarbeitung 2012. S. 96–103: TUDpress. Dresden (Studientexte zur Sprachkommunikation, 64), 2012.

[8] Dreuw, P.; Stein, D.; Deselaers, T.; Rybach, D.; Zahedi, M.; Bungeroth, J. and Ney H.: *Spoken language processing techniques for sign language recognition and translation.* In: Technology and Disability 20 (2), S. 121–133, 2008.

[9] Cooper, H.; Holt, B.; Bowden, R.: *Sign Language Recognition.* In: Thomas B. Moeslund, Adrian Hilton, Volker Krüger und Leonid Sigal (Hg.): Visual Analysis of Humans, S. 539–562: Springer London. London, 2011.

[10] Stein, Daniel, 2012: *Soft features for statistical machine translation of spoken and signed languages.* Dissertation. Aachen, Techn. Hochsch, Aachen, 2012.

[11] Wolff, M.; Klimczak, P.; Lindemann, J.; Petersen, C.; Römer, R. und Zoglauer, T.: *Die Kognitive Heizung.* In: Rüdiger Hoffmann (Hg.): Elektronische Sprachsignalverarbeitung 2014; TUDpress. Dresden, 2014.

[12] Huber, M.; Kölbl, C.; Lorenz, R.; Römer, R.; Wirsching, G.: *Semantische Dialogmodellierung mit gewichteten Merkmal-Werte-Relationen.* In: Rüdiger Hoffmann (Hg.): Elektronische Sprachsignalverarbeitung 2009; TUDpress. Dresden, 2009.

[13] Kölbl, C.; Huber, M.; Wirsching, G.: *Endliche gewichtete Transduktoren als semantischer Träger.* In: Bernd J. Kröger und Peter Birkholz (Hg.): Elektronische Sprachsignalverarbeitung 2011; TUDpress. Dresden, 2011.

[14] Mohri, Mehryar: *Weighted Automata Algorithms.* In: Manfred Droste (Hg.): Handbook of Weighted Automata, S. 213–254: Springer. Berlin (Monographs in Theoretical Computer Science), 2009.

[15] Lorenz, Robert; Huber, Markus: *Realising the translation of utterances into meanings by Petri Net Transducers.* In: Petra Wagner (Hg.): Elektronische Sprachsignalverarbeitung 2013. S. 103–110: TUDpress. Dresden (Studientexte zur Sprachkommunikation, 65), 2013.

[16] Huber, M.; Kölbl, C.; Lorenz, R.; Wirsching, G.: *Konstruktion von UMP-Transduktoren aus Wizard-of-Oz Daten.* In: Petra Wagner (Hg.): Elektronische Sprachsignalverarbeitung 2013. S. 111 – 118; TUDpress. Dresden, 2013.

[17] Wirsching, Günther; Wolff, Matthias: *Semantische Dekodierung von Sprachsignalen am Beispiel einer Mikrofonfeldsteuerung.* In: Rüdiger Hoffmann (Hg.): Elektronische Sprachsignalverarbeitung 2014; TUDpress. Dresden, 2014.