

KONSTRUKTION EINER MIKROLOKALEN GRAMMATIK MIT OPENFST AM BEISPIEL EINER HOME-ENTERTAINMENT-ANWENDUNG

Dennis Ritter und Günther Wirsching

*Katholische Universität Eichstätt-Ingolstadt
guenther.wirsching@ku.de*

Kurzfassung: Bei der Konstruktion eines Sprachdialogsystems ist es sinnvoll, die erforderlichen Sprachmodelle von vorne herein bedeutungsorientiert zu gestalten. Ein wichtiger Schritt ist dabei die Konstruktion mikrolokaler Grammatiken. An einem ausgewählten Beispielsatz aus einem WOz-Experiment zu einer Home-Entertainment Anwendung wird ein halbautomatisches Verfahren zum Erzeugen einer mikrolokalen Grammatik vorgestellt, wobei speziell hierfür entwickelte Softwarewerkzeuge zum Einsatz kommen.

1 Einleitung

Im Fokus dieser Arbeit steht die Modellierung der Syntax-Semantik-Schnittstelle im Rahmen des UASR (*Unified Approach for Speech Synthesis and Recognition*), das an der TU Dresden von Rüdiger Hoffmann et al. entwickelt wurde. Sowohl im Analyse- als auch im Synthesezweig des UASR soll diese Schnittstelle aus bidirektional verwendbaren *mikrolokalen Grammatiken* aufgebaut werden. Wir konzentrieren uns hier auf den Analysezwweig und zeigen die Konstruktion eines *Transduktors* (FST = *Finite State Transducer*), der einerseits eine Menge von zu einem *semantischen Schema* gehörender Wortfolgen beschreibt, und andererseits jeder solchen Wortfolge eine Darstellung ihrer Bedeutung zuordnet. Dazu wurden Software-Werkzeuge entwickelt, die eine Anbindung an OpenFST [1], und damit ans UASR, ermöglichen.

Wir beginnen mit formalen Definitionen grundlegender Begriffe unserer Semantikmodellierung, insbesondere (*kombinierbare*) *semantische Anker*, *Äußerung-Bedeutung-Paar* (UMP = *Utterance-Meaning Pair*), *semantisches Schema*, und *mikrolokale Grammatik* [4]. Dabei zeigt sich, dass bei praktischen Anwendungen *Merkmal-Werte-Relationen* eine wichtige Rolle spielen. Anschließend demonstrieren wir, wie solche Strukturen zur Analyse von Wizard of Oz-Daten eingesetzt werden können. Daraus ergibt sich ein Verfahren zur Generierung mikrolokaler Grammatiken aus geeignet annotierten Wizard of Oz-Daten. Zur Demonstration dieses Verfahrens haben wir Software-Werkzeuge entwickelt, deren Einsatz an einem Beispiel gezeigt werden soll.

2 Modellierung der Semantik

2.1 Formaler Hintergrund

Hintergrund der hier verwendeten Modellierung der Semantik ist das in [4] beschriebene allgemeine Semantikmodell. Formal basiert dieses auf der Zermelo-Fraenkelschen Axiomatisierung

der Mengenlehre, wobei es praktisch ist, der ursprünglichen Idee von Ernst Zermelo zu folgen und *Urelemente* miteinzubeziehen. Unsere Semantiksprache ist also ZFU (Zermelo-Fraenkel mit Urelementen); unter einem *ZFU-Objekt* verstehen wir entweder ein Urelement oder eine Menge in ZFU. Ein *semantischer Anker* ist, per definitionem, ein ZFU-Objekt, das sich (subjektiv) auf Realität bezieht. Dabei heißen zwei oder mehrere semantische Anker *kompatibel*, wenn sie (wieder subjektiv) zwei oder mehrere Aspekte der gleichen Realität referenzieren. Unter einer *Bedeutung* verstehen wir eine Menge kompatibler semantischer Anker. Die Bausteine einer natürlichen Sprache sind dann Äußerung-Bedeutung-Paare, bestehend aus einer Äußerung, die in eine Situation passt, wo auch die Bedeutung passt.

2.2 Anwendung auf automatische Dialogsysteme

Aus der subjektiven Sicht eines automatischen Dialogsystems bestehen alle möglichen Realitätsreferenzen aus Daten, die in einer geeigneten Datenbank vorliegen. Eine naheliegende Methode ist, solche Daten als *Merkmal-Wert-Paare*, z. B.

Nachname → Grant,

oder als Folgen von Merkmal-Wert-Paaren, z. B.

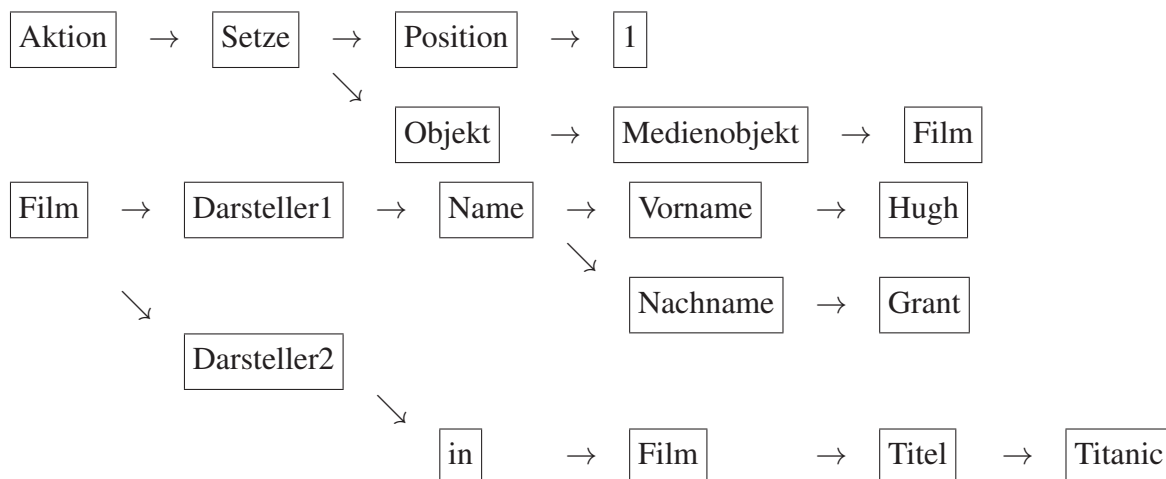
Medienobjekt → Film → Titel → Titanic,

darzustellen. Es ist klar, dass sich derartige Strukturen als ZFU-Objekte betrachten lassen: die Bezeichner „Nachname“, „Grant“, etc., sind die Urelemente, und Folgen sind mengentheoretisch konstruierte *n*-Tupel. Eine Bedeutung einer Äußerung besteht dann aus mehreren solchen Merkmal-Wert-Folgen, die insgesamt in vielen Fällen eine Baumstruktur ergeben.

Nehmen wir als Beispiel eine Home-Entertainment-Anwendung, wie sie in [3] und [2] betrachtet wurde, und betrachten wir die folgende Äußerung, die in dem in [3] beschriebenen Wizard of Oz-Experiment aufgenommen worden ist:

Auf Platz 1 soll der Film den ich letztens gesehen habe, mit äh Hugh Grant und äh der Schauspielerin die auch in Titanic mitgespielt hat.

Im Hinblick auf die Datenbank der Home-Entertainment-Anwendung könnte man dieser Äußerung die folgenden Merkmal-Wert-Strukturen als Bedeutung zuordnen:



Prinzipiell wäre es möglich, noch mehr Information aus der Äußerung in die Semantik abzubilden, z. B. das Geschlecht von „Darsteller2“, oder eine Referenz auf Daten, die mit der

Die Idee ist nun, in einer solchen Merkmal-Wert-Struktur bestimmte semantische Anker durch *Platzhalter* zu ersetzen – im Beispiel ist es naheliegend, den Namen des Schauspielers „*Hugh Grant*“ und den Filmtitel „*Titanic*“ offen zu lassen. Durch das Einfügen von Platzhaltern anstelle bestimmter semantischer Anker ergibt sich ein *semantisches Schema*. Zu jedem Platzhalter gehört eine Menge möglicher semantischer Anker, die an dieser Stelle sinnvoll eingesetzt werden können – im Beispiel wären das verschiedene Namen von Schauspieler oder Schauspielerinnen, bzw. verschiedene Filmtitel.

Ein Ziel dieser Arbeit ist die Konstruktion einer auf der Äußerung aus dem WOz-Experiment basierenden *mikrolokalen Grammatik*, die alle Varianten dieser Äußerung beschreibt, die sich durch Befüllen der Platzhalter mit verschiedenen passenden Datenbankeinträgen ergeben. Außerdem soll eine Methode realisiert werden, die jedem Satz der mikrolokalen Grammatik seine Bedeutung als Merkmal-Wert-Struktur zuordnet.

```
FVR[  
    Aktion[  
        Setze[  
            Position[  
                1  
            ] ][  
                Objekt[  
                    Medienobjekt[  
                        Film[  
                            Darsteller1[  
                                Name[  
                                    Vorname[  
                                        Hugh  
                                    ] ][  
                                        Nachname[  
                                            Grant  
                                        ] ][  
                                    ] ][  
                                    ] ][  
                                Darsteller2[  
                                    in[  
                                        Film[  
                                            Titel[  
                                                Titanic
```

3 Konstruktion des FST

Ein erster – naheliegender – Zugang zur Konstruktion einer mikrolokalen Grammatik wäre, sich einfach selbst ein paar Sätze zu überlegen, die die gewünschte Bedeutung ausdrücken, und aus diesen dann per Hand die Grammatik zu basteln. Bei einem solchen Vorgehen würden wir von einem individuellen Sprachverständnis ausgehen, das ausreicht, den gewünschten Ausschnitt einer natürlichen Sprache angemessen zu approximieren.

Besser ist jedenfalls ein Zugang, der sich auf Messwerte aus realitätsnahen Situationen stützt. Nun ist es schwierig, in einem Wizard of Oz-Experiment wirklich realitätsnahe Situationen zu erzeugen. Die Methode mit den semantischen Ankern ermöglicht es aber, wenigstens diejenigen Aspekte zu realisieren, die für die Bedeutung von Äußerungen aus Sicht des Dialogsystems relevant sind. Die im WOz-Experiment aufgenommenen Äußerungen sind dann als psychologische Messungen zu behandeln. Unsere FSTs werden so konstruiert, dass jede aufgenommene Äußerung als Anhaltspunkt für eine mikrolokale Grammatik genommen wird.

3.1 Syntaktische Aufgaben des FST

Prinzipiell hat der zu konstruierende FST zwei syntaktische Aufgaben:

1. Abstraktion von Datenwerten, und
2. Anpassungen im Kontext, um die grammatische Stimmigkeit der Äußerung zu erhalten.

Ausgehend vom vorliegenden Beispielsatz soll die mikrolokale Grammatik außer dem Beispielsatz selbst noch Sätze zulassen, bei denen die folgenden Bestandteile variiert sind:

- Anstelle des Schauspielernamens *Hugh Grant* sollen auch andere Namen von Schauspielern oder Schauspielerinnen verstanden werden.
- Statt „... *der Schauspielerin, die ...*“ soll auch „... *dem Schauspieler, der ...*“ verstanden werden können.
- Auch andere Filmtitel als *Titanic* sind zulässig.

Die Abstraktion von Datenwerten erfolgt dadurch, dass die zugehörigen Worte in der Äußerung zunächst lokalisiert und dann durch geeignete Platzhalter ersetzt werden. Im Beispiel betrifft das den Schauspielernamen und den Filmtitel.

Zur Behandlung von Kontextanpassungen kommt ein Markierungssystem zum Einsatz. Im Beispiel betrifft das die Ersetzung des Worts *Schauspielerin* durch *Schauspieler*. Der erste Schritt ist die Ersetzung des Worts *Schauspielerin* durch einen Platzhalter „_den_actor“. Eine Liste möglicher Einträge für diesen Platzhalter muss zu jedem Eintrag einen Hinweis auf sein grammatisches Geschlecht enthalten:

Schauspielerin	f
Schauspieler	m
Darstellerin	f
Darsteller	m

Die Anpassung der bestimmten Artikel erfolgt durch eine Ersetzungstabelle, die folgende Gestalt hat:

der	die	das
ein	eine	ein
welcher	welche	welches
dessen	deren	dessen
dem	der	dem
den	die	das
dieser	diese	dieses
dieses	dieser	dieses
diesem	dieser	diesem
diesen	diese	dieses
sein	ihr	sein
seine	ihre	seine
seines	ihres	seines
seiner	ihrer	seiner
seinem	ihrem	seinem
seinen	ihren	seinen
er	sie	es
ihn	sie	es
ihm	ihr	ihm

Hier wird ausgenutzt, dass in der deutschen Sprache die maskulinen Formen der angeführten Artikel und Pronomen so stark flektieren, dass sich eine zusätzliche Markierung des grammatischen Kasus erübrigt. – Für eine Anwendung in anderen Sprachen muss die Gestalt der Markierungen angepasst und ggf. erweitert werden.

Damit die Anpassungen bei der Konstruktion des FST wirksam werden können, müssen die anzupassenden Worte in der Ausgangsausßerung markiert werden. Das Format ist:

(anzupassendes Wort in der maskulinen Form)<(beeinflussendes Merkmal).

Mit dem „beeinflussenden Merkmal“ ist hier der Platzhalter für denjenigen semantischen Anker gemeint, durch dessen grammatisches Geschlecht die Form des anzupassenden Wortes bestimmt wird. Z. B. wird die Wortfolge „ der Schauspielerin, die “ zur Zeichenkette

```
dem<_den_actor   _den_actor   der<_den_actor
```

wobei das Komma weggelassen werden kann.

3.2 Der halbautomatische Schritt in der FST-Konstruktion

Die Konstruktion des gesuchten FST erfolgt, indem zunächst in einem halbautomatischen Verfahren die Semantik hinzugefügt und die Abstraktion der Datenwerte und die Kontextanpassungen vorbereitet werden. Zunächst wird die Äußerung von Satzzeichen befreit und automatisch tokenisiert, und das Ergebnis im openFST-Format mit zunächst leeren Ausgabesymbolen abgelegt. Diesen Schritt erledigt bei uns das PerlSkript

```
string2fst.pl
```

Als nächstes wird zu diesem Ergebnis die Bedeutung von Hand hinzugefügt – dabei ist es sinnvoll, darauf zu achten, dass die auszugebenen Bestandteile der Semantik möglichst genau an den Worten stehen, denen man die jeweiligen Bestandteile zuordnen möchte. Das ermöglicht im Prinzip auch ein Übertragen von Konfidenzen einzelner Worte auf Semantikbestandteile.

In unserem Beispiel ist das Resultat nach diesem halbautomatischen Schritt das folgende:

0	1	–	FVR[
1	2	Auf	Aktion[
2	3	Platz	Setze[Position[
3	4	1	]]]
4	5	soll	Objekt[
5	6	der	–
6	7	Film	Medienobjekt[Film[
7	8	den	–
8	9	ich	–
9	10	letztens	–
10	11	gesehen	–
11	12	habe	–
12	13	mit	Darsteller1[
13	14	äh	–
14	15	_first_name	Name[Vorname[_first_name]] [
15	16	_last_name	Nachname[_last_name]]
16	17	und	–
17	18	äh	–
18	19	dem<_den_actor	–
19	20	_den_actor	Darsteller2[
20	21	der<_den_actor	–
21	22	auch	–
22	23	in	in[
23	24	_film	Film[_film]
24	25	mitgespielt	–
25	26	hat	]]]]]]]]]
26			

Die erste Zeile ist immer: 0<TAB>1<TAB>–<TAB>FVR[. Die Nummer des Zustands, der als akzeptierend gekennzeichnet werden soll, wird alleine in eine eigene Zeile geschrieben. Eine Datei mit diesem Inhalt ist der Ausgangspunkt für die weitere – automatisierte – Verarbeitung.

3.3 Der automatische Teil der FST-Konstruktion

Die automatische Generierung der mikrolokalen Grammatik benötigt folgende Daten:

1. Ein annotierten und markierten Ausgangssatz.
2. Zu jedem Platzhalter eine Liste der in Frage kommenden semantischen Anker.
3. Eine Ersetzungstabelle für die Kontextanpassung.

Unser Softwarewerkzeug erzeugt dann zunächst alle nach diesen Vorgaben möglichen Sätze. Anschließend wird die Grammatik durch Verbaumen etwas kompakter gestaltet. Abb. 1 zeigt,

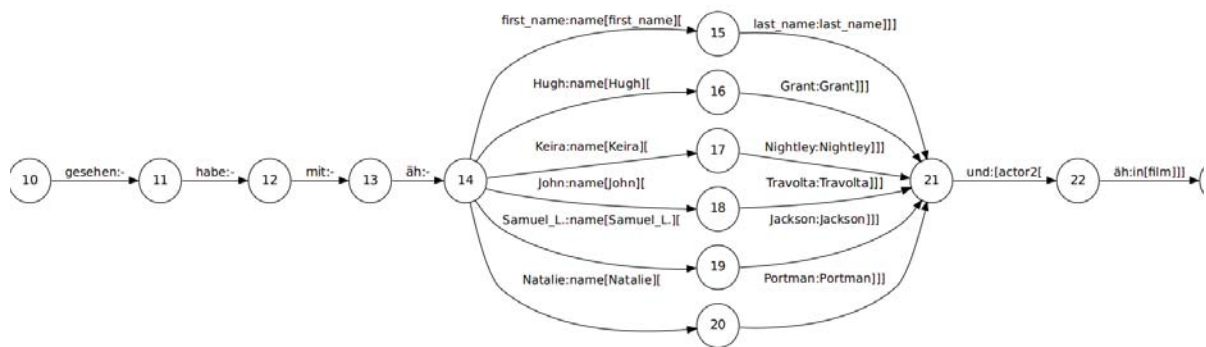


Abbildung 1 - Die Variationsmöglichkeiten bei den Namen erzeugt ein „Knäuel“ im FST.

wie durch Einfügen von Varianten ein FST, der eine mikrolokale Grammatik darstellt, verändert wird.

4 Schlussbemerkung und Ausblick

Die vorliegende Arbeit stellt einen ersten Schritt zur praktischen Erzeugung mikrolokaler Grammatiken aus einzelnen Sätzen aus Wizard of Oz-Experimenten dar, indem zunächst alle Kombinationen aus gegebenen Listen semantischer Anker zugelassen werden. In unserem Beispiel in Abb. 1 haben wir nur solche Vorname-Nachname Kombinationen zugelassen, die in einer Liste von Schauspielernamen stehen. Es bleibt eine Frage des Dialog-Designs, ob und ggf. in wie weit unzulässige Kombinationen – wie z. B. nicht zusammenpassende Vor- und Nachnamen – in der Erkennergammatik zugelassen werden sollen – in dieser Richtung ist unser Softwarewerkzeug recht flexibel einsetzbar.

Ein Fernziel wäre es, mit der Zeit eine Datenbank mikrolokaler Grammatiken aufzubauen, die dann in unterschiedlichen Sprachdialogsystem eingesetzt werden könnten.

Literatur

- [1] *OpenFST*. <http://www.openfst.org/twiki/bin/view/FST/WebHome>.
- [2] HUBER, M., C. KÖLBL, R. LORENZ und G. WIRSCHING: *Konstruktion von UMP-Transduktoren aus Wizard-of-Oz Daten*. In: *ESSV2013*, S. 111–118.
- [3] KARNAGEL, K., R. RÖMER, S. ROGGE und J. LINDEMANN: *Semantische Modellierung von System-Benutzer-Interaktionen am Beispiel einer Home-Entertainment Steuerung*. In: WOLFF, M. (Hrsg.): *ESSV2012*, Bd. 64 d. Reihe *Studentexte zur Sprachkommunikation*, S. 306–313, Dresden, August 2012. TUDPress.
- [4] WIRSCHING, G. und R. LORENZ: *Towards meaning-oriented language modeling*. In: *IEEE 4th Int. Conf. on Cognitive InfoCommunication, Budapest, Hungary, December 2-5, 2013*.
- [5] WIRSCHING, G. und M. WOLFF: *Semantische Dekodierung von Sprachsignalen am Beispiel einer Mikrofonfeldsteuerung*. In diesem Tagungsband.