Adjunct Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ,13), October 28–30, 2013, Eindhoven, The Netherlands.

Development of a Conversational Speech Interface Using Linguistic Grammars

Hansjörg Hofmann, Ute Ehrlich, Sven Reichel, André Berton
Daimler AG, Ulm, Germany
{hansjoerg.hofmann, ute.ehrlich, sven.reichel, andre.berton}@daimler.com

ABSTRACT

In the automotive environment, intuitive in-car speech interfaces are crucial in order to reduce driver distraction. The design of an intuitive speech interface poses a great challenge to the speech recognition and natural language understanding component of a speech dialog system since human language allows speakers to create an infinite number of sentences. In this paper, a linguistic grammar approach, which incorporates linguistic knowledge in the grammar design in order to develop flexible grammars, is presented. Based on this approach a conversational speech dialog system, which allows German users for booking a hotel has been developed.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces

1. INTRODUCTION

Todays' in-car speech dialog systems (SDS) are command-based. Due to the novel in-car Internet access, the number of accessible applications in the car increases quickly. The cognitive load needed in order to control the SDS will rise with the number of available applications as the number of possible commands will also increase. Hence, a conversational and intuitive speech interface is necessary to ease voice-control and to reduce driver distraction. Apple's Siri, for example, allows users to speak with their mobile phone as if they would talk to a human being. However, usually, only the first step in the human-machine interaction can be performed by speech. After the first spoken utterance the user has to continue with haptic input to achieve his goal. In the automotive environment, a turn-by-turn conversational speech interface is targeted in order to allow the driver to keep hands on the wheel and eyes on the road.

The design of a conversational speech interface poses great challenges to the automatic speech recognition (ASR) and natural language understanding (NLU) modules of an SDS since "any speaker of a human language can produce [...] an infinite number of sentences"[1]. The human spoken language allows speakers to create sentences with the same meaning in many possible ways by e.g., using synonyms, reordering constituents or even concatenating phrases.

There are two main approaches to model the user's language for ASR[4]. Statistical language models (SLM) provide statistical information on word sequences and are used to predict the next word in a sentence. However, in order to generate a reliable language model, a huge amount of data is needed to train the SLM. In addition, there is no efficient way to semantically link the output of the SLMs with the NLU. The second approach is based on grammars, which use rules to model permissible word sequences. The design of a wide and flexible grammar is time-consuming and often legal word sequences, which were not anticipated, are ruled out or syntactical erroneous sentences are falsely accepted. However, grammar rules can simply be extended to deliver a semantic annotation of the recognized word sequence.

AutomotiveUI'13, October 28–30 2013, Eindhoven, Netherlands.
Adjunct Proceedings

In this paper, a linguistic grammar approach, which incorporates linguistic knowledge in the grammar design, is presented. The ASR and the NLU component use separate grammars which are both based on the same linguistic grammar sources. In the next Section, the linguistic grammar concept is described. Based on this approach we developed a conversational SDS, which allows German users to book a hotel, which requires a wide and flexible grammar due to the various input parameters. The SDS prototype is described in Section 3 and finally, conclusions are drawn.

2. LINGUISTIC GRAMMAR CONCEPT

The linguistic grammar approach is based on several lexica of words and a set of syntax rules which is illustrated in Figure 1.

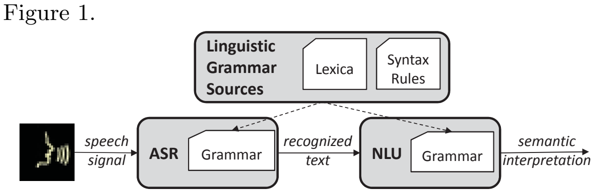

Figure 1: ***Linguistic Grammar Concept.***

There are lexica for each lexical category (verbs, determiners, nouns, etc.). Each lexicon contains a set of words and possible synonyms. Furthermore, each word in the lexica is labelled with its semantic meaning and its morphological properties. E.g., for each entry in the verb lexicon the verb form, the tense, the mode, the number and the person are specified (see Table 1).

The syntax rules describe how the words in the lexica can be concatenated to syntactic categories. For each combination, the lexical category and its required morphological properties are specified. A concatenation is valid if the morphological properties of the different words match. Thereby, it is ensured that only syntactical phrases are generated. Sample syntax rules are illustrated in Table 2.

The lexica and the syntax rules should be defined in a general way in order to reuse these components for future gram-

Table 1: *Sample Verb Lexica.*

Lexical Entry	Semantics	Verb Form	Tense	Mode	No.	Person
suchen (search)	search	infinitive				
suche (search)	search	finite	present	ind.	sg	first
suche (search)	search	finite	present	imp.	sg	first
finden (find)	search	infinitive				

mars. The lingustic grammar sources are employed to generate grammars for the language model of the ASR module and the keyphrase spotting technology of the NLU module. The specification of the different grammars and their employment are explained in the following.

Table 2: *Sample Syntax Rules.*

Syntact. Cat.	Ex.	Lex. Cat.	Case	No.	Gend.
NP (Noun Phrase)	*ein Hotel (a hotel)*	Determiner	nom	sg	neutr
		Noun	nom	sg	neutr
NP (Noun Phrase)	*die Hotels (the hotels)*	Determiner	nom	pl	fem
		Noun	nom	pl	fem
PP (Prepositional Phrase)	*in einem Hotel (in a hotel)*	Preposition	dat		
		Determiner	dat	sg	masc
		Noun	dat	sg	neutr

2.1 ASR Grammar

Based on the lexica and the syntax rules the grammar of the ASR can be specified. The desired syntax rules and the required semantic values of the constituents have to be indicated in order to specify a grammar rule:

$search= *V[search] + NP[theNumber, roomType] + PP[in, starHotel]*

By indicating the semantic value of the constituents only synonyms with the same meaning are selected. The syntax rules help to generate all possible syntactical sentences. The above-mentioned sample rule produces numerous sentences with the same meaning, for example (English: "Search for 2 double rooms in a 4-star hotel"):

"Suche 2 Doppelzimmer in einem 4-Sterne-Hotel."

"Suche 2 Doppelzimmer in einem Hotel mit 4 Sternen."

"Finde 2 Doppelzimmer in einem 4-Sterne-Hotel."

Complex use cases like a hotel booking require the user to indicate multiple search parameters in order to retrieve a list of hotels. These search parameters can occur in different orders and in different combinations which still have to be taken into consideration and have to be specified. However, the use of syntax rules reduces this problem to a permutation of constituents which can be generated automatically.

2.2 NLU Grammar

In order to interpret the recognized utterance, the NLU component uses phrase spotting techniques based on the lexica entries and the syntax rules. A sample rule for interpreting the hotel category is illustrated in the following:

$hotelCategory = *PP[in, starHotel] | NP[starHotel] |N[starHotel]*

By applying the above-mentioned phrase spotting rule all phrases like "in einem 4-Sterne-Hotel", "ein 4-Sterne-Hotel", "4-Sterne-Hotel" and possible synonym phrases are interpreted. The NLU phrase spotting is independent from the employed recognizer engine. Thus, the engine can be replaced and the SDS stays flexible towards new recognizer technologies.

3. SDS PROTOTYPE

Based on this linguistic grammar approach we developed a conversational SDS, which allows German users to book a hotel. The design of the SDS's grammar is based on speech data we collected in an online user study[2]. For ASR we employ Nuance's VoCon®3200[1] embedded speech recognizer. A graphical user interface has been designed in order to support the speech dialog.

In order to evaluate our approach we compare the performance of a grammar-based ASR engine based on our approach with SLM-based ASR engines. Furthermore, we investigate the respective interpretation result when sending the ASR output to the NLU module based on our grammar approach. The grammar-based ASR engine is Nuance's VoCon®3200[3] embedded speech recognizer. In addition, we access offboard Nuance's Dragon NaturallySpeaking[2] and WebSearch server[4] which both employ domain-unspecific SLMs. In a preliminary lab experiment we collected 177 conversational utterances with 24 people (m/w=15/9, average age=29) driving in a simulator using our prototype at an early development stage. We use this corpus to evaluate the performance of the ASR engines on word accuracy (WA) [3]. Furthermore, the concept accuracy (CA) [3] of the NLU module is assessed which is crucial to a successful SDS performance. The results are illustrated in Table 3.

Table 3: *Evaluation Results.*

Recognizer Engine	WA	CA
Grammar-Based VoCon	*81.3%*	72.6%
SLM-based Dragon	*88.4%*	72.4%
SLM-based WebSearch	*88.7%*	75.5%

The results show that SLM-based ASR performs better than grammar-based which is due to out-of-vocabulary words missing in the lexica. Training of the SLMs on the specific domain would further improve the WA. Concerning CA, all setups achieve similar results whereof the WebSearch recognizer performs best since its ASR result contains the most semantically relevant constituents. Despite the decreased ASR performance, with our grammar-based approach these constituents, which are crucial for successful NLU processing, are recognized.

4. CONCLUSIONS

In this paper, we presented a linguistic grammar approach, which incorporates linguistic knowledge in the grammar design of an SDS. The ASR and the NLU component use separate grammars which are both based on the same linguistic grammar sources. Based on this approach a conversational SDS, which allows German users for booking a hotel has been developed. The ASR and NLU performance results proved our concept and the flexibility of our approach.

Within the scope of the EU funding project GetHomeSafe the SDS prototype is evaluated in driving simulation studies on usability and its impact on driving performance.

5. REFERENCES

[1] V. Fromkin, R. Rodman, and N. Hygams. *An Introduction To Language - Seventh Edition*. Cengage Learning, 2003.

[2] H. Hofmann, U. Ehrlich, A. Berton, and W. Minker. Speech interaction with the internet - a user study. In *Proceedings of Intelligent Environments*, Guanajuato, 2012.

[3] ITU-T Suppl. 24 to P-Series Rec.,. Parameters describing the interaction with spoken dialogue systems. International Telecommunication Union, Geneva, 2005.

[4] M. F. McTear. *Spoken Dialogue Technology: Towards the Conversational User Interface*. Springer, London, 2004.

[1] http://www.nuance.com

[2] http://www.nuancemobiledeveloper.com