

# Natural language processing in dialogue systems with spoken input

Claus Povlsen  
København

## Abstract

This paper describes the linguistic analysis done within a research project, the goal of which is the development a spoken language understanding system. The first part outlines the overall system design including a description of the constraints to be dealt with in systems handling spoken input. The second part contains a description of how the domain-specific sublanguage was defined and how domain-knowledge was exploited in the syntactic and semantic analyses in order to make the linguistic description as precise and unambiguous as possible.

## 1 Introduction

The project *Spoken Language Dialogue Systems*<sup>1</sup> aims to produce two functioning prototypes of spoken language understanding systems which will enable a given user to make flight reservations automatically via the telephone. The task for the Centre for Language Technology in the project is to define and implement the natural language user interface between the speech recognition and the knowledge based components of the system.

Machine-based speech understanding systems differ profoundly from more "traditional" NLP systems which usually take input in a textual format. Processing speech to a symbolic representation involves a huge number of time consuming computations and has thereby set narrow bounds for the domain-specific sublanguage, i.e the user's linguistic interface to the developed system. Its interactive nature is another characteristic feature of the system which has influenced the choice of which type of sentence constructions were included in the grammatical coverage of the sublanguage. Finally as the project is application-oriented (near) real time performance is required, which has been decisive for the definition and delimitation of the lexical and grammatical coverage of the system.

---

<sup>1</sup>The basic research project is sponsored by the Danish Technical Research Council and is a cooperative effort between the Centre for Language Technology in Copenhagen, the Centre for Cognitive Informatics in Roskilde and the Centre for Speech Technology in Åborg (coordinating partner).

## 2 The overall system design

The overall architecture of the initial version of the system is modular, with communication between modules being handled by a Dialogue and Control manager, specially designed for graphically defining and executing task-oriented dialogues see Larsen et al. (1993) and Bækgaard et al. (1992).

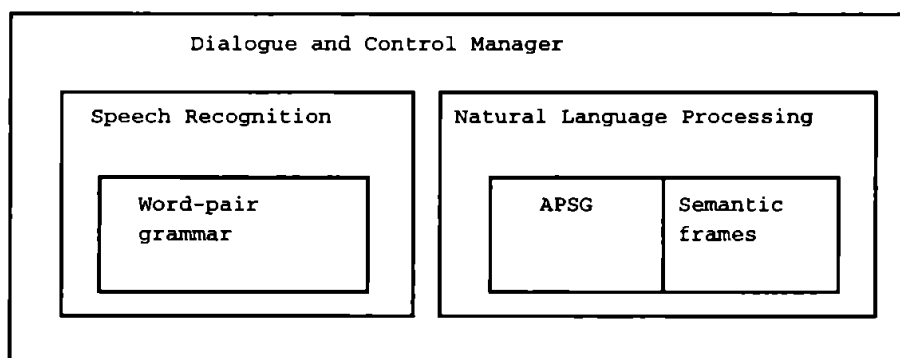


FIG 1: System architecture

The Dialogue and Control Manager receives sentence hypotheses from the speech recognizer, which in turn are sent to the parsing module for analysis. Each sentence hypothesis consists of a list of lexical references that are used by the parsing module for looking up the items in a lexicon. As a result of the NLP, the sentence-semantic information – formally expressed in framelike structures – is returned to the Dialogue and Control Manager for making decisions as to what actions should follow.

In the following, focus will first be on how the word-pair grammar used in speech recognition and the syntactic analysis of the NLP component interact. Thereafter the correlation between the syntactic analysis and the semantic interpretation will be described.

Without access to linguistic knowledge constraining possible combinations of the word forms in the vocabulary, the speech recogniser will in principle regard the number of legal word sequences as being equal to the number of word forms raised to the second power. In practically oriented spoken understanding systems in which real time performance is a decisive design criterion, quality speech recognition is only obtainable if the number of word models under consideration at any point in the acoustic processing is reduced. In order to reduce these reference patterns to be matched, a word-pair grammar, expressing knowledge of legal word-pairs in the sublanguage is defined. The word-pair grammar is automatically generated from the word sequences expressed in the unification based APSG (Augmented Phrase Structure Grammar). This ensures that the grammatical coverage of the APSG is a subset of the coverage of the word-pair grammars. This close relation between the

mentioned grammars has determined a precise and unambiguous formulation of the syntactic rules in the APSG.

In general the syntactic and semantic analyses can be completely separated or integrated in various ways. The graphic representation of the NLP component in the above figure is meant to illustrate that the syntactic and semantic descriptions are distinct, although during processing their applications are interwoven. This will be discussed further below.

### **3 Definition and delimitation of the domain-specific sublanguage**

Experience has shown difficulties implementing large systems with broad lexical and grammatical coverage. In light of this, attention has recently been directed to investigating the possibility of defining subsets of general languages in a principled fashion. Based on analyses of communication within delimited areas, research efforts have concluded that such domains of a general language are characterized by limited vocabularies and use of linguistic phenomena. This corresponds well with the realization that current speech recognition technology must set narrow limits on the linguistic coverage in spoken understanding systems.

In order to define the sublanguage within the application chosen, two different types of data have been collected. A travel agency (DanTransport) gave permission to tape on-site one hour of dialogues between their agents and clients. Furthermore several series of simulated man-machine dialogue experiments (Wizard of Oz-experiments) have been conducted. For a more detailed description see Dybkjær and Dybkjær (1993).

The relative importance of the different kinds of collected data depends on whether the dialogue structure in spoken dialogue systems are strictly system-directed or user-directed. If the dialogue structure is rather unconstrained and therefore comparable to dialogues between humans, recorded human-human data would be a significant source for delimiting the vocabulary. On the other hand, in directed dialogue mode, which constitutes a communication situation different from human-human dialogues, data from simulated human-machine experiments should form the basis for the definition of the domain-specific vocabulary. As in fact the system under development is highly system-directed, the data from the WoZ-experiments have constituted the primary source for defining the domain-specific sublanguage.

The sum of the clients' utterances from the simulated human-machine dialogues have been defined as constituting the domain-specific

sublanguage within the chosen application. Under due consideration to the limited speech recognition capacity, the defined sublanguage has been modified and extended based on acquired knowledge about the domain and about language in general. After adjustment of the vocabulary size, the number of word forms totalled 500.

After having investigated the collected corpus in order to register the linguistic phenomena, the most frequent used syntactic constructions were identified and included in the grammatical coverage. The most dominant phenomenon in the corpus was the presence of a large number of elliptic constructions. As the mode of the dialogue in the simulated system was system-directed, the wizard's (system's) authoritative way of asking questions made it unnecessary and irrelevant for the client to repeat the information already expressed, in which case he just added new information - linguistically expressed as a partial sentence, an ellipsis. A typical example is:

The wizard: *Po hvilket tidspunkt afgør det ønskede fly?*  
At what time does the desired plane depart?

The Client: Toogtyve femogfyrre  
10:45 PM

In general the implementation of elliptic constructions is implemented by allowing one constituent to be the single element in a sentence rule. In addition domain-specific constituents and word classes have been defined. The rules covering the client's utterance in the above example would thus be as follows:

S :- Hour\_p  
Hour\_p :- Card\_p

'Hour\_p' is a domain-specific non-terminal covering, among other things, cardinal phrases ('Card\_p') expressing *time* and constitutes a single element in a sentence rule. For a more detailed description of the syntactic analysis see Povlsen and Music (1993).

#### **4 Exploitation of domain-specific knowledge in the syntactic analysis**

As mentioned above the syntactic grammar in the system serves two purposes. Besides being used for making a structural analysis as a first step towards a semantic interpretation of the utterance, it also forms the basis for the automatically generated word-pair grammars applied for reducing the number of word models during the acoustic signal

processing. The latter function requires strict formulation of the grammar for syntactic analysis in order to make the linguistic knowledge used in speech recognition as precise and constrained as possible.

Besides a reduction in quantity (the number of words and syntactic constructions used), semantic restrictions in sublanguages make it possible to analyse the utterances in a much more specific and precise way. For example on the basis of an analysis of the main verbs with focus on their contextual word patterns domain-specific selectional restrictions and categories can be determined and exploited in the linguistic processing.

#### 4.1 Selectional restrictions

Selectional restrictions express constraints on combinations of lexical units in a given context. By focussing on the argument structure for lexical units, interrelated semantic concepts can be identified. How domain-specific knowledge is expressed in selectional restrictions can be illustrated by looking at following sample utterances from the collected corpus.

*jeg vil bestille en billet*  
I want to order a ticket  
*jeg vil lave en reservation*  
I want to make a reservation

If only syntactic restrictions are defined in a sentence rule covering these sample utterance i.e. SUBJECT AUX VERB OBJECT, it would permit acceptance of a large number of meaningless combinations. Consider for instance the following example:

\* *en billet vil bestille et barn*  
A ticket wants to order a child

If the activation of the syntactic rules is not constrained further, the coverage of the generated word-pair grammars will end up being too broad (loose) and will thereby exceed the limits for the permitted number of active word models, leading to poor speech recognition results. As a first step towards solving this problem, all the domain-specific main verbs and their contextual patterns were examined in concordances. Based on this analysis the domain-specific concepts which link lexical units together were identified. The outcome of the investigation concerning *bestille* and *lave* was that they had the following concept valency frames:

arg1[+HUMAN], arg2[+TICKET [+CONCRETE]]

arg1[+HUMAN], arg2[+TICKET [-CONCRETE]]

The knowledge of domain-specific selectional restrictions is expressed in the lexicon and grammar of the NLP module. The implemented parsing algorithm for the syntactic analysis is an Earley-based, left-right chart parser. This means that in the initial phase of the processing all lexical information is assigned to the input words in the chart and then used by the syntactic grammar as well-formedness constraints in further processing. As mentioned above, the constraint contribute to a lowering of the perplexity in the automatically generated word-pair grammars, thereby improving speech recognition quality.

## 4.2 Domain-specific categories

In practically oriented NLP systems covering subsets of general languages, definition of word classes and structures specific for delimited subject areas is a method often applied for making the overall process more efficient.

For instance in the domain of flight reservation, it will be more straightforward to define domain-specific categories for phrases such as *klokken toogtyve femogfyrre* (at twenty-two forty-five), *halv ni* (half past eight), which do not fit the ordinary patterns of nominal phrases. By defining non-terminals for these constructions the syntactic component of the system will be activating fewer syntactic rules and thereby reducing the number of wrong word sequences in the coverage of the automatically generated word-pair grammars.

Based on the utterances from the collected corpora, interchangeable elements are assigned with common values in defined attribute value pairs. The cardinals ranging from one to ten is thus coded with: 'minut\_fg1=yes' in the lexicon, while all other cardinals are coded with 'minut\_fg1=no'. Using this restriction in the grammar rules covering the following phrases:

*klokken fem minutter i halv tolv*  
o'clock five minutes to half two  
at one twenty-five

*klokken otte minutter over halv fire*  
o'clock eight minutes past half two  
at three thirty-eight

prevents the acceptance of "wrong" word sequences such as:

\* *klokken elleve minutter i halv tolv*  
o'clock eleven minutes to half twelve  
at eleven nineteen

\* *klokken seksten minutter over halv fem*  
o'clock sixteen minutes past half five  
at four forty-six

Dependency among word sequences in the corpora concerning *Hour* is expressed by defining non-terminals. In the sample utterances the word forms within each sequence *fem minutter i halv* and *otte minutter over halv* are dependent on each other in that all the words must be present in order to express a meaningful utterance. In the syntactic grammar these sequences are expressed as follows:

```
half_p =  
  [  
    {cat=card_p, minut_fg1=yes},  
    {cat=n, dalu=minut},  
    {cat=p},  
    {cat=adj, lex=halv}  
  ].
```

By writing a syntactic grammar that is domain-specific, the overall NLP will be more efficient and makes possible an effective matching strategy for elliptic constructions

## **5 Exploitation of domain knowledge in the semantic interpretation**

As mentioned above, there are several linguistic advantages to handling sublanguage instead of general language. Besides a reduction in quantity, semantic restrictions in special languages make it possible to describe sublanguages in a more specific and precise way. Besides reduced polysemy and the possibility of exploiting domain-specific selectional restrictions and defining domain-specific categories, adequate interpretation of an utterance is also simpler, the number of semantic roles in the domain being easier to grasp.

Based on knowledge of known goals within the domain and on a linguistic analysis of the collected corpus, domain-relevant semantic roles have been defined. Based on the above mentioned identification of domain-specific constituents and parts of speech, word classes deemed as essential

for fulfilment of domain-specific goals of the system were then expressed in the semantic representation either as 'frames' or as subordinated 'slots'. Head concepts such as *Reserve* (filled by *bestille/reservere*) is thus assigned a 'frame', while 'hour' is represented as a 'slot' in this 'frame'.

key: Reserve

slots:

```
persons
id_number
date{ }
hour()
from
to
```

As mentioned in the description of the overall system design, the syntactic and semantic rules of the NLP-module are separate. However, before processing the two types of descriptions are compiled into the same format. Whenever a new syntactic constituent is found in the input, the parser immediately checks whether any of the semantic rules can be applied. This means that the system searches for semantic interpretations as early as possible in the parse process without waiting for the syntactic analysis to finish.

Error recovery is done on ungrammatical speech recognition results. If the input-utterance *jeg vil bestille en billet til Ålborg* (I would like to order a ticket to Ålborg) is recognised as *jeg vil bestille en Billund til Ålborg* (I would like to order a Billund to Ålborg) the flexible parsing design of the NLP-module makes it possible to interpret the utterance adequately. Despite the fact that the input to the NLP-component is out-of-coverage (no sentence rule exists for this word sequence) access to the semantic information assigned the subconstituents during the processing will be sufficient for making an adequate semantic representation. This is done by implementing a mechanism which collects the results stored in the chart and gather them together under a "dummy" sentence symbol.

## 6 Conclusion

Satisfactory speech recognition results presuppose access to linguistic knowledge during the acoustic processing. A much debated subject concerning speech understanding systems has been how to integrate speech recognition and NLP in order to simultaneously achieve optimum recognition quality and to generate an adequate syntactic analysis. For a discussion see Brøndsted (1993) and Povlsen and Music (1992). Focus here has been on describing how knowledge of the domain-specific sublanguage has been exploited in the natural language processing in the



initial prototype of the system. This is done partly by application of selectional restrictions and domain-specific categories in order to reduce the generality of the structural analysis and partly by delimitation of the space of interpretation in the semantic analysis.

## References

- Brøndsted, Tom. 1993. *Integration af akustisk genkendelse og natursprogsprocessering*. Datalogivistik foreningen Årsmøde 1993.
- Bækgaard, Anders, A. Roman and P. Wetzel. 1992. *Advanced Dialogue Design -- DDL Tool and ICM*. ESPRIT SUNSTAR 2094, Deliverable IV. 6–2.
- Dybkjær, Laila and Hans Dybkjær. 1993. *Wizard of Oz Experiments in the Development of the Dialogue Model for P1. Spoken Language Dialogue Systems* – project report 3.
- Larsen, Lars Bo., T. Brøndsted, H. Dybkjær, L. Dybkjær and B. Music. 1993. *Overall Specification and Architecture of P1. Spoken Language Dialogue Systems* – project report 2.
- Povlsen, Claus and Bradley Music. 1992. *Natursprogsprocessering i dialogsystemer med talt input*. SKRIFTEN PÅ SKÅRMEN, no. 6. HHÅ.
- Povlsen, Claus and Bradley Music. 1993. *Sublanguage definition and specification for P1. Spoken Language Dialogue Systems* – project report 4 (in progress).