The Role of Semantics in Spoken Dialogue Translation Systems

Scott McGlashan*

University of Saarbrücken, Germany
scott.mcglashan@coli.uni-sb.de

Abstract

In this paper, we consider the role of semantics in the spoken dialogue translation systems. We begin by looking at some of the key properties of an existing spoken dialogue system, namely the SUNDIAL system which provides flight and train information over the telephone, and how these properties affect the design methodology and functionality of spoken translation systems. These properties include the effects of speech processing, designing the system to meet the needs of users, and an analysis model which clearly separates the linguistic, conceptual, pragmatic and task levels. In this model many task functionalities are dependent upon, and sometimes realizable by, the semantic and pragmatic analysis components. Central to this approach, is the use of *underspecified* semantic representations which are further specified as and when required by domain and/or task analysis. This model can be applied in the development of spoken translation systems with two important effects: monolingual semantic and pragmatic analysis can be carried out by processes independent of, but correlated with, the (translation) function of the system; and the main functions of the transfer processes are to further specify the representations for the target language and to deal with mismatches between source and target language representations. We illustrate this approach with semantic analyses of German utterances required for translation in the VERBMOBIL spoken dialogue translation system.

1 Introduction

Since the early 60's machine translation has been seen as an application of Natural Language Processing (NLP) which has promised the development of translation systems useful for translators and naive users alike. While considerable progress has been made in the development of computer-based aids for the translator, such as terminological database systems, success in the development of (fully) automated translation systems has been less clear[1]. In comparison, a more recent application area of NLP, Spoken Language Systems (SLSs), have been successful in developing research prototypes and commercial systems which provide useful services to the general public (Church and Hovy, 1993). These systems ranged from

*This work was partly funded by the German Ministry for Research and Technology (BMFT) in the framework of the Verbmobil project under grant 01 IV 101 R. The responsibility for the contents of this study lies with the author.

[1] The obvious exceptions are systems which offer domain-restricted translation; for example, Taum-Meteo system for English-French translation of weather forecasts (Hutchins, 1986: 228-31).

simple, single digit/word voice response systems — allowing retrieval of answerphone messages, credit card balances, etc — to the more advanced class of interactive dialogue systems offering access to information services, such as flight information and reservation (Mangold, 1994). This raises the questions of what are the methodological and design principles of SLSs, and whether these principles can be applied to the design of machine translation systems. These questions are timely if the next decade is to see the development of spoken dialogue translation systems.

2 Spoken Dialogue Systems

For comparison with machine translation, one of the most interesting class of SLSs are spoken dialogue systems for information services such as train information, flight information and conference registration. They are interesting for two reasons.

Firstly, the complexity of the domain is comparable to that envisaged for many spoken translation systems. For example, SRI's Spoken Translator System, which provides English-Swedish translation over the telephone, operates in the ATIS flight information domain commonly used in spoken dialogue systems (Agnäs et al, 1994). Likewise, the VERBMOBIL system, discussed below, operates in the domain of appointment scheduling.

Secondly, these systems are *task-oriented* in the sense that user input is only analyzed to the depth required for task processing within a limited domain. In interactive dialogue systems, the appropriate levels of analysis are semantic and pragmatic: i.e. the task component exploits the (perceived) meaning and function of user utterances to determine the system response appropriate to the current stage in the dialogue. In the ASURA speech translation system, the translation regularities are similarly described in terms of semantic and pragmatic information (Morimoto et al., 1993).

The main principles of spoken dialogue systems can be found in the SUNDIAL system. The goal of the SUNDIAL project (Speech UNderstanding in DIALogue) was to build real-time integrated dialogue systems capable of maintaining co-operative dialogues with users over standard telephone lines (Peckham and Fraser, 1994). Systems have been developed for four languages – French, German, Italian and English – within the task domains of flight reservations and enquiries, and train enquiries. Here we shall consider three principles of this system and their implications for translation.

2.1 Speech Input

Analysis of spoken language faces an immediate problem: input is an acoustic signal which is more or less continuous and lacking in natural sub-divisions equivalent to words. For example, there will not be any significant acoustic break between the words when uttering *Is BA174 on time?*. In order to provide a representation for subsequent grammatical analysis, techniques have been developed for segmenting the acoustic signal, classify these segments into sub-word units ('feature extraction'), and then comparing these units with a database of known templates ('words') to identify those which best match the input.

The most common technique is the statistical technique of Hidden Markov Modelling which takes into account the effects of co-articulation, compensates for timing differences arising from differences in the speed at which words are spoken, and deals with differences between speakers (Holmes, 1988). This technique uses a large collection of speech data in order to learn statistical correlations between acoustic features and words. The output of this process

is a set of hypotheses consisting of representations of the identified words, their temporal positions, and their statistical likelihoods.

In addition to the basic problem of recognition, most spoken dialogue systems take as input *spontaneous* speech. In comparison to *read* speech, spontaneous speech introduces naturally occurring phenomena including:

extra-linguistic phenomena such as coughs, blows, lip 'smacks', and filled pauses; for example, *Can you erhm tell me ahhm about ....*

re-starts where the speaker reiterates part of an utterance after an interruption; *What is the depart- the arrival time of BA173?*

While these phenomena are gradually being included in the training processes, they still tend to constitute one of the principle stumbling blocks for speech recognition.

There are two important consequences of speech recognition for analysis in spoken dialogue systems. Firstly, recognition of spontaneous speech is currently only feasible with a vocabulary size of around 1500 words or less. While this may limit the linguistic coverage of the system, it does offer the advantage that a subsequent level of analysis can be restricted in scope: syntactic, semantic and pragmatic analysis need only model, or at least fully realize, phenomena within the scope of the speech recognition. Secondly, even with this limited vocabulary, speech recognition is not 100% accurate. For example, one of the SUNDIAL recognizers trained on collected spontaneous speech data yielded 74.4% word accuracy and 50.9% sentence accuracy (Baggia et al., 1994: 247). While accuracy is continuously improving — e.g. through the use of prosodic information, larger speech databases and word co-occurrences — subsequent processes still need to compensate for recognition performance.

2.2 User-Centered Design

In building spoken dialogue systems, it has long been recognized that they must be designed to have the functionalities expected by users. This requirement, however, leads to a 'chicken and egg' situation: how can system designers know how people will react to computers prior to the development of the system, and how can the system be developed prior to an understanding of users' behaviour and requirements?

The naive approach is to analyze human-human data in the same task domain. In the case of a dialogue system for providing flight information, the system is, in a sense, playing the role of a human information provider and so the behaviour users demonstrate in this situation can be used as a basis to develop the system's functionalities. However, while this type of analysis may tell us much about how humans interact, it will not *necessarily* reveal how people will interact with a computer-based system. Not only do these dialogues reveal a level of recognition (or even linguistic) performance far in advance of current technology, but *simulated* human-computer dialogues show that people interact with machines in a very different way[2].

Data collected during *Wizard of Oz* simulation experiments show that people use a different language and have different expectations about the capabilities of a computer-based information service (MacDermid, 1993). Simulation experiments involve a person pretending to be an intelligent computer. Human subjects are then led to believe that they will be interacting with an actual computer (either through screen-based exchanges, or though telephone mediated verbal exchanges) when in fact they are connected to the experimental

[2]More recently, the same result has been found for actual systems.

agent, suitably disguised, who is pretending to be the computer. The behaviour subjects display in interacting with systems differs, for example, in terms of:

vocabulary size: subjects use approximately 40% of the word types observed in human-human dialogues

sentence complexity: many sentence types, such as relative clauses, are of low frequency

referential domains: the domain typically consists of just the participants and the subject matter (no third person references)

error tolerance: subjects tolerate some (speech) errors if they are quickly 'repaired' by the system

On the basis of this type of evidence, it appears that users are prepared to adapt their interaction patterns to the limitations of spoken dialogue systems: they do not expect the system to provide all the communicative abilities of a human service provider so long as they obtain the information they require.

This 'adaption' phenomenon allows a spoken dialogue system to be developed with limited recognition and analysis capabilities, but still provide a useful service to the general public. This can be exploited in spoken dialogue translation systems: in a given domain, users will not necessarily expect the same quality of translation of a machine as they expect of a human translator. For example, a human translator may translate *Ja am Dienstag den sechsten April hätte ich noch einen Termin frei* as *I'm still free on Tuesday, sixth April* while the 'simpler' translation *OK on Tuesday the sixth of April I still have a slot free* may satisfy the user in an appointment scheduling domain. The empirically-based user-centered design strategy suggests that spoken translation systems should provide appropriate, consistent, and coherent translations rather than clever translations which mimic complex human capabilities[3].

2.3 Domain-oriented Analysis

In spoken dialogue systems like SUNDIAL, analysis of user utterances is *domain-oriented*. At the semantic and pragmatic levels, user utterances are analyzed relative to a discourse model built up during the dialogue (Heisterkamp et al., 1992). Semantic analysis establishes which domain objects are being talked about and how they have changed relative to the previous discourse context. At the pragmatic level, the semantic analysis is used in conjunction with a dialogue model to determine the illocutionary effect of the utterance; for example, whether the utterance is informing the system of new or changed information about the flight, confirming information given by the system, or requesting information. Finally, at the task level, the semantic and pragmatic analyses are used to determine whether the user has provided sufficient information for retrieving a solution from the database, or whether further information is required.

2.3.1 Multi-level Analysis Model

Central to this approach is an underlying analysis model which clearly separates the language-specific from the language-general, domain-independent interpretation from

[3] In fact, there is a well-known danger of 'faking' more complex analyses: the user may attribute too much linguistic competence to the system and when the user finds that the system is not consistently capable of demonstrating it, they are dissatisfied.

domain-dependent, and each are separated from the task dependent aspects (Eckert and McGlashan, 1993). This model is motivated by the aim of designing *generic* analysis components which can be customized for language, domain and task. In the SUNDIAL project, the same components were used for English, French, German and Italian languages, domains such as train and flight timetables, and the (information retrieval) tasks of enquiry and reservation. Of course, for commercial exploitation these distinctions can be collapsed so as to establish a direct connection between the syntactic analysis of an utterance and its domain-specific and task-specific interpretation (Magadur et al., 1993).

2.3.2 Semantic Analysis

Semantic and pragmatic analyses are central to the determination of task interpretation. Syntactic analysis of an utterance is treated as a 'stepping-stone' which allows the construction of semantic and pragmatic representations. These latter representations abstract away from surface-specific realizations while preserving features of information structuring and sequencing indicated by syntax[4]. One of the main influences on the semantic representation is **underspecification**. If all semantic information required for task interpretation were to be specified in the lexicon, then the number of lexical entries would be enormous: a different lexical entry would potentially be required for each contextually-dependent sense. Instead, the lexical semantics of an expression contains information which is shared between these different contexts and, guided by the discourse and dialogue contexts, further information is added to their interpretation. This approach can be modelled with an inheritance-based conceptual type hierarchy and a set of rules for refining the conceptual types (McGlashan, 1993).

The hierarchy consists of a set of *typed* concepts partially ordered in terms of subsumption. Concepts can be atomic or complex. For example, *six* may be associated with an atomic concept type NUMBER with the value 6. Complex concept types each contain one or more roles, where the value of the role is itself a concept type. The roles each describe the relationship between the main concept type and the concept type fulfilling the role. For example, as shown in Figure 20-1, the type GO is specified with *theme* and *goal* roles; the *goal* role expresses a relation between the 'going' event and the intended location of the *theme*. Since the hierarchy is inheritance-based, types subsumed by GO, such as DRIVE, are not only more specific types of a 'going' event but also inherit its roles.

The type hierarchy is partitioned into linguistically-oriented concept types and domain-oriented concept types. The former types can be directly realized in natural language; for example, GO and DATE. The latter types are domain-specific types, such as DBFLIGHT and DBTRAIN, which are closely connected with task-level objects. Semantic interpretation proceeds by constructing a domain-independent representation in terms of linguistically-oriented conceptual types; integrating (or 'anchoring') the conceptual representation into the discourse context established from the interpretation of previous utterances; and finally, linking the conceptual representation to a domain-dependent task representation. Crucial to these processes are necessary and default inference rules. Necessary inference rules describe relations between conceptual types which always hold in a given situation; default rules are only contingently valid for a given situation. For example,

(1) CLOCKTIME: 24hourtime(necessary, <hour value>, <minute value>, <am-pm value>, <24hour value>)

(2) CLOCKTIME: minutetime(default, <minute value>)

[4] For example, the active-passive distinction can be represented in terms of relative informational prominence of semantic arguments (Kay et al., 1994: 94).

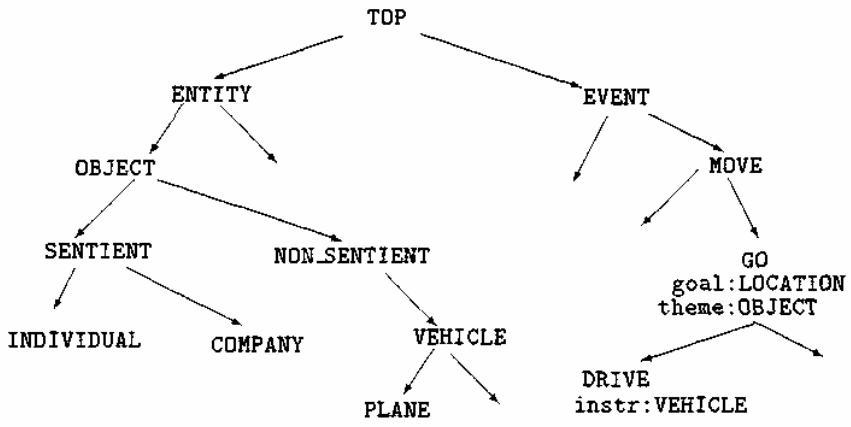


Figure 20 - 1 A simple Conceptual Type Hierarchy

(3) DEPART: equality(necessary, <place>, <journey departure place>)

(4) JOURNEY: equality(necessary, <departure place city value>, <dbflight source city>)

where the rule in (1) describes a necessary relationship between 24 hour clocktime and the 'am-pm' distinction in English (*10 30 pm = 22 30 hours*), the rule in (2) describes a relation which only holds in the absence of more specific information; e.g. the user may not mention the minutes when describing the time — *I want a flight leaving at 10.* The rule in (3) describes a necessary relationship in the travel domain: i.e. the departure place in a leaving event is also the departure place in a journey. Finally, the rule in (4) establishes a relationship between the departure city in a journey concept with the sourcecity parameter in a domain-specific concept.

Two aspects of the semantic interpretation process need to be mentioned. The first concerns defeasibility: i.e. situations where there is a 'conflict' between semantic information and where the conflict is systematically resolved through the 'defeat' of one property in favour of the other. In the normal situation, the discourse model is monotonically extended on the basis of information from the user and the application of necessary inference rules. However, there are situations where the discourse model needs to be non-monotonically extended. For example, a default rule has been applied and subsequently the user provides a different value from that given by the rule. This case can be simply dealt with by systematically giving priority to user-supplied information.

A more serious problem in spoken dialogue systems arises due to the uncertainty of speech recognition[5]. Consider, for example, the following dialogue fragment:

(5) User: I want to fly from Paris.
System: Do you want to fly from Paris?
User: Not Paris, I want to fly from Perros.

where the system confirmation reveals that *Perros* has been misrecognised as *Paris*. This type of 'necessary' defeasibility can be dealt with by treating the discourse model as a

[5]The problem can also arise naturally since speakers can be inconsistent, or merely careless, in giving information.

partial view of the world (i.e. not making the 'closed world assumption') and allowing the content of the discourse model to be seen from different *views* where consistency need only be maintained within a given view. This approach is illustrated in Figure 20-2.

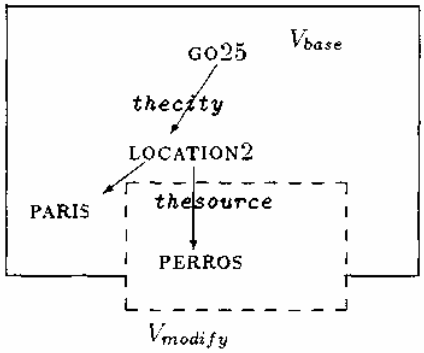


Figure 20 - 2 Different views in the Discourse Model

2.3.3 Pragmatic Analysis

Another important aspect of the interpretation process is its relationship with pragmatic analysis. Pragmatic analysis in spoken dialogue systems is concerned with determining the impact of utterances in terms of dialogue structure. The structure of information service dialogues can be described in terms of *exchanges*, *interventions*, and *dialogue acts* (Bilange, 1991). Figure 20-3 illustrates this type of analysis. The utterances are structured into ex-

$$E_0 \left[\begin{array}{ll} E_1 & \left[\begin{array}{ll} I_{A_1} : request & When\ would\ you\ like\ to\ leave? \\ R_{B_1} : inform & November\ the\ 13th \\ Ev_{A_2^1} : confirm & November\ the\ 13th \end{array} \right. \\ E_2 & \left[I_{A_2^2} : request \quad At\ what\ time\ would\ you\ like\ to\ leave? \right. \end{array} \right.$$

Figure 20 - 3 Dialogue Structure

changes: the exchange E_0 consists of two subexchanges, E_1 and E_2, which are concerned with establishing the departure date and time respectively. Each exchange is initiated (I), reacted to (R) and, optionally, evaluated (Ev). Within each part of an exchange, the utterances are assigned one or more dialogue acts which describe what the utterance is intended to achieve (Bunt, 1989). Speaker A opens exchange E_1 with a question (I_{A_1}). Speaker B reacts with *November the 13th* (R_{B_1}) which is analyzed as an *inform* act since (a) its semantic analysis indicates that it is providing new (date) information about a domain object (flight) and (b) this information can be treated as an informative reaction to the preceding question. Speaker A (positively) evaluates this exchange with a repetition ($Ev_{A_2^1}$), which is categorized as a *confirm* act since the semantic analysis indicates it is old information, and opens a new exchange with a request for a departure time ($I_{A_2^2}$)[6]. Since the interpretation of

[6]Note these functions are implicit in the utterance. Speakers can, of course, be explicit about the purpose of their utterance by, for example, using a performative verb *I want to know when you would like to leave*

an utterance is, in part, determined by its location in the dialogue structure, this can affect its translation; for example, *Yoroshiku onegaishi masu* would be translated into English as *thank you* in a confirmation phase of the dialogue, but as *good bye* in a closing phase.

2.3.4 Implications for Transfer

This analysis model presented in the preceding sections has been developed for information service dialogues on the assumption that it could be easily extended to other tasks. The questions we briefly address here are how can it be extended to the translation task and what are the implications for the transfer model.

One feature of the information service task is that it involves *information extraction*: i.e. linguistic semantic information is extracted from user utterances so as to build a domain-level representation and, from that representation, information is extracted to build a representation suitable for database access. This has allowed us to use a simple, linguistic semantic representation for utterances which neutralizes some subtle distinctions within and between languages. For example, the verbs *go* and *gehen* are assigned the same conceptual type GO. This treatment can be justified on the grounds of corpus analysis where the expressions are consistently used with the same domain-specific sense. Furthermore, the inference rules mapping between the linguistic semantics and domain-specific semantics allowed us to treat domain conceptual types as 'instances' of the linguistic conceptual types and also as transformations of them. For example, the rules mapping the GO linguistic concept into the domain type JOURNEY can be seen as conceptual transformations motivated by domain-specific knowledge.

For dialogue translation systems, the relation between the domain and task levels is not directly relevant[7]. Rather the transfer relation is defined between utterances in different languages at the level of domain-oriented semantic representations. It follows from our model that the transfer relation can involve both instantiation and transformation of these representations: i.e. in some cases, the target representation is simply an extension of source representation, while in others it is a transformation. If this is the case, then our model of transfer can be constrained by (a) identifying the set of transformation relations, and (b) defining a 'defeasibility' hierarchy which predicts the relative priority of information types — semantic, pragmatic, register — when there is a mismatch.

This view of translation can be seen as a compromise between interlingua- and transfer-based approaches. It follows the interlingua position in treating the transfer relation as primarily based on semantic (or conceptual) information but does not follow it to the extent of assuming the same concepts necessarily underlie translation-equivalent utterances in different languages. Rather, for a given domain, it admits that conceptual structures appropriate in the source language may need to be transformed into conceptual structures appropriate for the target language: "Translation is not meaning preserving" (Kay et al., 1994: 85). On the other hand, while following the transfer approach of using rules to map between source and target structures, it differs in that (a) syntactic structure is not (directly) used to define the relation and (b) pragmatic, as well as semantic, information is used to define the relationship.

Finally, this approach can be seen as a 'distributed' approach to translation. Rather than localize or encapsulate all translation-relevant functions in a single component — the transfer module — some of these functionalities can be carried out by other components. For example, the semantic component can infer domain-specific interpretations of user input, so

(Ripplinger, 1994).

[7] unless, of course, the system also functions as an information provider. Usually, this function is performed by the person whom the translation is being provided for.

obviating the need for 'disambiguation' transfer rules[8]. In pursuing this approach, it should become clear to us which functions really need to be carried out by the transfer component and which are more appropriate for other components in spoken dialogue systems. This approach is being explored within the context of the VERBMOBIL project.

3 The Verbmobil System

The VERBMOBIL project combines speech technology with machine translation techniques in order to develop a system for translation in face-to-face dialogues (Wahlster, 1993). The VERBMOBIL system will provide English translation for negotiation of business appointments between German and Japanese users who have only a passive knowledge of English[9]. The major requirement is to provide translation as and when users need it, and do so in real-time. In order to meet this requirement, the system is composed of time-limited processing components which perform acoustic, syntactic, semantic and pragmatic analysis, dialogue management as well as generation and synthesis.

The user, when requiring translation, activates the VERBMOBIL device and speaks in German. The speech recognition component processes this input and produces a word lattice representing hypotheses about what was said[10]. The parsing component processes this lattice so as to determine which sequences are well-formed with respect to its grammar. The output is a sequence of syntactic representations. For each syntactic representation, a linguistically-oriented semantic representation is constructed and evaluated so as to disambiguate expressions, assign dialogue acts and update the current discourse and dialogue models. The dialogue component, on the basis of the assigned dialogue act(s), predicts what type of utterance might follow so as to guide speech recognition[11]. The transfer component takes the semantic and pragmatic analysis of the input and builds a semantic and pragmatic representation for the target language expression; the transfer component exploits semantic and pragmatic information in order to map between source and target languages (Maier and McGlashan, 1994). The generator then constructs a syntactic structure for the target sign suitable for synthesis in English.

Three aspects of the VERBMOBIL scenario distinguish it from the typical scenario of spoken dialogue systems. The first is that the system plays the role of a dialogue *mediator* rather than a dialogue *partner*. The dialogue is principally between the two people rather than a person and a machine. On occasions when German speakers are unable to express themselves directly in English, the system acts on their behalf by translating their German utterances. Secondly, one of the effects of this dialogue mediation role is that the contextual information available to the system is incomplete. In its standard role as a dialogue partner, the system has full access to the context: i.e. all utterances are either spoken by the system, or direct towards it. In the VERBMOBIL scenario, however, that part of the dialogue in English is not directly towards it. Without this 'English' context, processes which rely on contextual analysis, must be able to operate without access to the full dialogue context. Although techniques for identifying key phrases in the English — such as dates, time as well as positive and negative evaluation phrases — will be used, it not yet clear how robust and accurate they will be. Finally, since users may be non-native speakers, translation strategy and quality will need to reflect their level of competence. Translations, especially idiomatic translations such as *How about October?* for *Wie wäre es im Oktober?*, which may be very

[8] for example, lexical transfer rules that map a lexeme in one language into different lexemes in another language depending upon the contextually-appropriate sense.

[9] In the first phase of the project, the system is limited to German-English translation.

[10] Prosodic information can also be represented in the lattice.

[11] In cases where no analysis has been assigned, the dialogue component initiates a clarification dialogue. For example, if the user spoke too loudly, an appropriate utterance in German is formulated and synthesized.

'natural' to a native speaker, may be beyond the competence of a non-native speaker. More serious are translations which depend upon cultural knowledge; for example, the translation of *Allerheiligen* as *All Saints Day*[12].

3.1 Semantic Representation

It has long been recognized in formal semantics that semantic formalisms which are denotationally interpreted, declarative and compositional have considerable methodological advantages over formalisms which are not. For example, formalisms without a denotational interpretation, such as the one used in the SUNDIAL system, lack a sound theoretical basis for controlling inferential and resolution processes. The problem until quite recently has been that the representation which satisfies these criteria (Montague's Intensional Logic) is not capable of providing semantic interpretation beyond the sentence level. However, a new class of 'dynamic' semantic formalisms, including Discourse Representation Theory, has been developed to address this issue (Kamp and Reyle, 1993). DRT characterizes meaning in terms of discourse representation structures (DRSs), which are interpretable within first-order, model-theoretic semantics at the level of propositions, and crucially provides mechanisms to model context-dependent interpretation and context-change. The semantic formalism in VERBMOBIL, λ-DRT, is a compositional version of DRT augmented with conceptual information, and pays special attention to phenomena characteristic of spoken dialogues (Bos et al., 1994).

3.1.1 Compositionality

λ-DRT combines the basic features of DRT with Montague-style Extended Type Theory to obtain compositionality. In essence, DRSs (Discourse Representation Structures – pairs consisting of a set of discourse markers and a set of conditions on these markers) are taken as the basic meaning expressions but λ-abstraction over DRSs allows the construction of complex meaning expressions. This approach allows the bottom-up compositional construction of semantic representation from syntactic structures[13]. For example, the representation of *vorschlagen* in (6) indicates the verb is involved in two composition operations (indicated by 'λ'): it combines with the 'object' semantics as argument and functor respectively (λt); and then it combines with the 'subject' semantics in the same relationship (λi).

(6) $\lambda t.\lambda i.$ $\boxed{\begin{array}{l} e \\ vorschlagen(e, \{i_{agent}, t_{theme}\}) \end{array}}$

3.1.2 Integrating Conceptual Information

The semantic formalism has been extended to incorporate conceptual information similar to that found in the SUNDIAL semantics. Thematic roles explicitly label the relationship between two discourse markers; in (6), the *agent* labels relationship between i and e. The

[12]Note that this cannot always be solved by simply being more precise and translating it as *1st November* since this produces vacuous translations for confirmatory utterances — *Ist Allerheiligen der erste November?* (*Is the 1st November the 1st November?*).

[13]It also allows the underspecification of quantifier scope since resolution of scope ambiguities is not obviously relevant for translation in a limited domain.

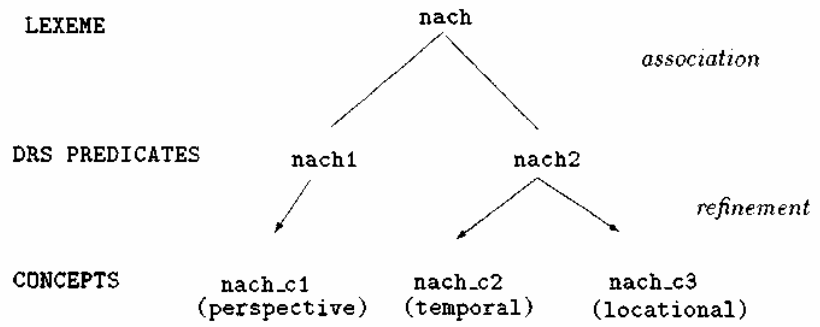


Figure 20 - 4 Lexeme-Semantic-Concept Relationships

discourse markers themselves have been assigned sorts, such as e for event, i for individual and t for time, which correspond to linguistically-oriented conceptual types[14]. Like the conceptual type hierarchy sketched in Section 2.3.2, the domain model in VERBMOBIL can express both necessary and default relations between conceptual types and has the advantage that it has a well-defined interpretation in Descriptive Logic (Hoppe et al., 1993). The semantic representation has also been augmented with 'hooks' to *instantiated* conceptual representations in the discourse model. For example, when the representation in (6) combines with the semantics of a subject and object, and the resulting representation is evaluated against the domain context model, the event marker e will be extended to e_{inst10} indicating that it corresponds to $inst10$ in the model. The effect of this is that the result of global conceptual refinement, which may occur for example during anaphora resolution, is accessible from the semantic representation upon which transfer operates.

Unlike in SUNDIAL, there is no direct relationship between lexeme and concept. Each lexical item contains one or more semantic representations. There is more than one representation if either of the lexical expressions are assigned multiple syntactic categories, or the senses can only be described using different semantic structures. Otherwise, they are given the same semantic representation, independent of whether they differ at the conceptual level. The relationship between lexeme, semantic and conceptual representation is illustrated in Figure 20-4. The preposition *nach* is given two DRS representations: *nach1* is used for the 'perspectival' sense (e.g. *nach meinem Terminkalender - according to my diary*); while *nach2* subsumes the temporal and spatial senses (*nach Berlin - to Berlin* and *nach dem Mittagessen - after lunch*). The conceptual level describes each of these senses.

Access to the conceptual level permits ambiguities to be resolved prior to transfer. Some of these ambiguities are domain-independent and can be resolved using local context. For example,

(7) Ich komme mit dem Auto
(I am coming **by** car)

(8) Ich komme mit meine Frau
(I am coming **with** my wife)

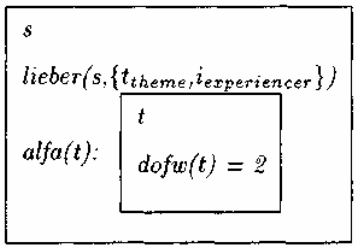
where the preposition fulfils either an instrumental or concomitant role depending on the

[14] Sorts and thematic roles are not defined in the formalism, but in a domain model representing linguistic and conceptual information relevant to the application domain.

conceptual type of its local NP argument. Other ambiguities are domain-dependent and non-local. For example, *bei mir* is ambiguous in many utterances between a location and perspectival reading and consequently can be translated as *at my place* or *for me*. In the domain of appointment scheduling, this ambiguity can be resolved if the context maintains a domain object, comparable with DBFLIGHT in SUNDIAL, describing the current appointment and whether both participants have agreed to it. Prior to an agreed appointment being fixed, *bei mir* is refined to its perspectival reading. Once it has been agreed, then, unless local context tells us otherwise, it is refined to its location reading.

However, contextually-driven conceptual refinement is not always necessary for transfer, or always possible. The predicates used in the DRSs split into 'pivot' interlingua-style predicates and language-specific predicates. Transfer can operate on interlingua-style semantic predicates without considering how they are refined at the conceptual level. For example, the predicate *dofw(t) = 2* (second day of week), as shown in (9) below, is the representation for both *Dienstag* and *Tuesday*. In other cases, the DRS representation neutralizes syntactic structure mismatches, such as head switching, which typically require complex transfer rules. Neutralization occurs when the same predicate argument structure is maintained in both source and target languages. Since the DRS semantics adopts a Davidson treatment of adjuncts, the syntactic distinction between a complement and adjunct is neutralized at the semantic level. For example, *Dienstag ist mir lieber* has the following representation:

(9)



< i, speaker>

where the adjunct *mir* ($i_{experiencer}$) is incorporated into the argument structure of the predicate. From this representation, two English translations can be given: *I prefer Tuesday*; or more literally, and maintaining the relative information prominence of the arguments, *Tuesday is preferable for me*. Finally, more complex cases concerning scope resolution — such as the apparent movement of negation from the predicate in *verwählen* to its implicit argument in *to dial the wrong number* — still await investigation.

3.1.3 Anaphora, Ellipsis and Modality

The formalism pays special attention to three types of phenomena frequently occurring in appointment dialogues: anaphoric, elliptical and modal expressions.

(10) Ich schlage den Dienstag vor
I propose Tuesday

(11) Das paßt echt schlecht bei mir
That is really impossible for me

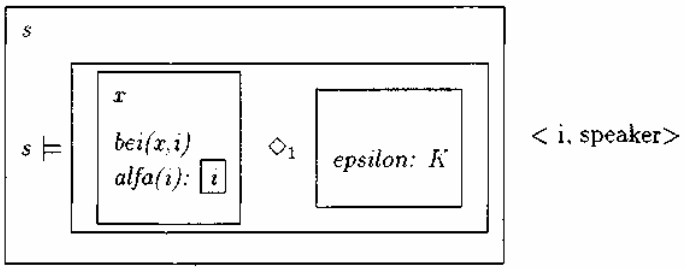
In (10) *den Dienstag* is an anaphoric definite description, and in (11) *paßt* introduces a modality while *Das* is elliptical in referring to a previously established proposition.

The anaphoric status of expressions is represented using an 'alfa' condition which explicitly indicates that the discourse marker should be evaluated against the discourse model to

determine its antecedent. For example, the representation for *Dienstag* in (9) indicates that the discourse marker t in the 'alfa' structure should be linked to a previously established discourse marker. If there is no suitable marker, it is 'accommodated' in the discourse model: i.e. a new discourse object is introduced. Resolving anaphoric expressions is not only useful for generation — it has the *choice* of whether to use a full or reduced description — but also for transfer. For example, the German verb *belegen* may be translated as *reserve*, *book* or *take* depending on whether the conceptual type of the object is a *building*, *time* or *course*, respectively. If the object of the verb were a pronoun, then anaphora resolution could supply a conceptual type of the appropriate specificity. Of course, given the potential lack of context in VERBMOBIL, then this set of (necessary) refinement rules needs to be augmented with a 'default' rule for cases where the conceptual type is insufficiently specific for transfer or generation.

In the VERBMOBIL domain, once participants have established a conversational topic, they are not explicit about it in every utterance. In arranging a date for a meeting, participants may use elliptical expressions, such as *das*, to refer to abstract entities constructed from earlier established events or propositions ('<participants>') meeting on Tuesday'). Elliptical expressions are represented by 'epsilon' conditions as shown in the representation for *Das paßt echt schlecht bei mir* below:

(12)



where *das* is represented by 'epsilon: K' (K indicates that the ellipsis refers to a proposition). Like alfa conditions, the resolution of elliptical expressions can contribute to sense refinement for transfer. In addition, some classes of elliptical expressions need to be resolved for the generation and domain-specific default resolutions may need to be specified; for example, by default *das* refers to a proposition describing a meeting on some proposed, but unknown, date.

Modal expressions, such as *paßt*, are represented using the $\Diamond$ structure (indicating possibility), as in standard DRT, but with additional components necessary for spoken dialogue data. In (12), the DRS to the right of $\Diamond$ indicates the argument of the modal, while the DRS to its left indicates the 'perspective' upon which the modality is based ('for my point of view'). In addition, the subscript on $\Diamond$ represents a scalar feature acting as an intensifier or weakener of the possibility. The scalar value is determined by the semantics of adjectives such as *gut* and *schlecht*: the default value is 3, and the combined effects of *echt* and *schlecht* reduce this to 1.

Since modal expressions have little propositional content, they are prime candidates for translation which is oriented more to the pragmatic level of analysis than the semantic. For in addition to the dialogue act discussed in Section 2.3.3, the pragmatic level in VERBMOBIL also includes *dact_arg* (the argument of the dialogue act), *tone* (the degree of suitability) and *perspective* (dialogue participant) features. So, assuming the representation in (12) is assigned the dialogue act 'reject', the following pragmatic representation can be assigned:

(13)
```
[dact: reject
 dact_arg: epsilon: K
 tone: strong
 perspective: speaker]
```

This allows a number of different utterances, or even a 'stereotypical' utterance, to be generated in the target language, each of which preserves the communicative intent of the speaker.

3.1.4 Fragmentary Input

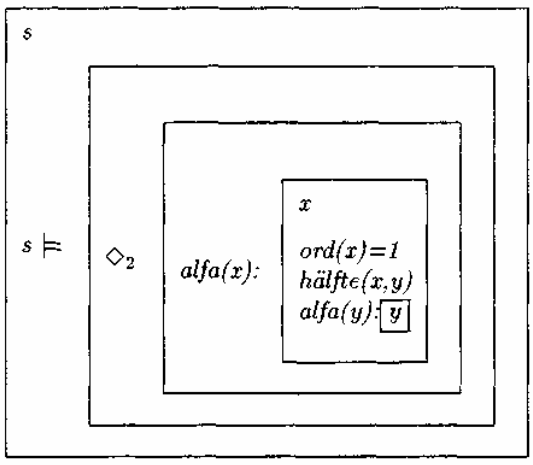
Unlike conventional text-based systems, which assume the sentence as the maximal domain of processing, speech-based systems have the 'turn' as their domain of local analysis. In the VERBMOBIL scenario, a 'turn' can be defined as the period which begins when the 'translate' button is pressed and ends when it is released. During that time, the user may utter a number of sentences, or phrases, separated by (arbitrarily long) pauses. In addition, segmentation of the utterance into multiple phrases may also arise on account of the speech processing component: even when the speaker produces a single sentence, the speech component may output a lattice whose best hypothesis corresponds to a sequence of grammatical 'fragments'. In both cases, the semantic analysis component has as input a sequence of representations which are not integrated at the syntactic level. For example,

(14) [Können wir den Oktober vergessen] [aber] [nicht] [den November]
(*We can forget October, but not November*)

(15) [Die erste Hälfte] [Das ist schlecht]
(*The first half. That is bad.*)

We are pursuing the approach that many of these fragments can be related at the semantic level: i.e. the semantic analysis yields a unitary representation (Heisterkamp et al., 1992). In (14) the fragments cannot be related at the syntactic level since *nicht* is a verb modifier, and so cannot modify *den November*, and *aber* requires two constituents of the same syntactic type. These fragments can be related at the semantic level, however, if the latter fragments are treated as elliptical and the status of *aber* as a contrastive discourse connective is used to determine the missing semantic information. In (15) the fragments can be integrated into a unitary semantic representation by resolving the anaphora *das* with the semantics of *die erste Hälfte* as shown in (16)[15]:

(16)



[15]Note that this resolution only requires local context.

where, prior to resolution, the DRS to the right of $\diamond$ consists of an empty alfa expression. In both cases, the unified semantic representation offers the transfer and generation components more translation options; for example, the unitary representation of (14) can be realized more explicitly as *We can forget October, but we should not forget November* or less explicitly as *We should not forget November*; and (16) can be realized as *The first half is bad*. By providing a semantically-based treatment of fragmentation in spoken dialogue, this approach allows transfer to adopt a 'reduction-oriented' translation strategy as used by professional translators and interpreters.

4 Conclusions

In this paper we have explored machine translation from the perspective of spoken dialogue systems, focusing on the role of semantics. Three principles of dialogue systems, illustrated with the SUNDIAL system, were discussed and related to the translation task. Recognition of the limitations of speech recognition, the importance of appropriate empirical analysis, and the power of a semantic and pragmatic analysis model appropriate to the domain, can contribute to the development of spoken dialogue translation systems. The semantic representation used in the VERBMOBIL system was described, compared with the approach to semantics in SUNDIAL, and shown to be useful for a semantically- and pragmatically-oriented approach to translation. The approach has been implemented in a 'mini' demonstration system and will be extended to cover a more extensive fragment of German. By adopting some of the principles of spoken dialogue systems, we expect that the VERBMOBIL project will result in spoken translation systems which provide a practical service to the general public on the basis of a theoretically sound analysis model.

References

Agnäs et al, M-S. (1994). Spoken language translator: First-year report. Technical report. SRI International. Cambridge, England.

Baggia, P., E. Gerbino, E. Giachin, and C. Rullent (1994). Experiences of spontaneous speech interaction with a dialogue system. In Nieman, H., R. de Mori, and G. Hanrieder (eds) *Progress and Prospects of Speech Research and Technology*. Sankt Augustin, Germany: Infix. 241–8.

Bilange, E. (1991). A task independent oral dialogue model. In *Proceedings of the 5th Annual Meeting of the European Chapter of the Association for Computational Linguistics*, Berlin. 83–88.

Bos, J., E. Mastenbroek, S. McGlashan, S. Millies, and M. Pinkal (1994). A compositional DRS-based formalism for NLP applications. In *Proceedings of the Internatonal Workshop on Computational Semantics, Utrecht*.

Bunt, H. C. (1989). Towards a dynamic interpretation theory of utterances in dialogue. In Elsendoorn, Ben A.G. and Herman Bouma (eds) *Working models of human perception*. New York: Academic Press.

Church, K. W. and E. H. Hovy (1993). Good applications for crummy machine translation. *Machine Translation* 8. 239–58.

Eckert, W. and S. McGlashan (1993). Managing spoken dialogues for information services. In *Proceedings of the 3rd European Conference on Speech Communication and Technology*. 1653–6.

Heisterkamp, P., S. McGlashan, and N. J. Youd (1992). Dialogue semantics for spoken dialogue systems. In *Proceedings of the International Conference on Spoken Language Processing*. Banff, Canada.

Holmes, J. N. (1988). *Speech Synthesis and Recognition*. Wokingham: Van Norstrand Reinhold.

Hoppe, Th., C. Kindermann, J.J. Quantz, A. Schmiedel, and M. Fischer (1993). BACK V5 tutorial and manual. KIT Report 100, Technical University of Berlin.

Hutchins, W. J. (1986). *Machine Translation: past, present and future.* Chichester: Ellis Horwood.

Kamp, H. and U. Reyle (1993). *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT.* Dordrecht: Kluwer.

Kay, M., J. M. Gawron, and P. Norvig (1994). *Verbmobil: A Translation System for Face-to-Face Dialog.* Number 33 in CSLI Lecture Notes. Standford: CSLI.

MacDermid, C. (1993). Features of naive callers' dialogues with a simulated speech understanding and dialogue system. In *Proceedings of EUROSPEECH'93*, Berlin, Germany.

Magadur, J-Y., F. Gavignet, F. Andry, and F. Charpentier (1993). A French oral dialogue system for flight reservations over the telephone. In *Proceedings of EUROSPEECH'93*, Berlin, Germany.

Maier, E. and S. McGlashan (1994). Semantic and dialogue processing in the Verbmobil spoken dialogue translation system. In Nieman, H., R. de Mori, and G. Hanrieder (eds) *Progress and Prospects of Speech Research and Technology.* Sankt Augustin, Germany: Infix. 270–3.

Mangold, H. (1994). Speech technology and telephone services. In Nieman, H., R. de Mori, and G. Hanrieder (eds) *Progress and Prospects of Speech Research and Technology.* Sankt Augustin, Germany: Infix. 212–19.

McGlashan, S. (1993). Heads and lexical semantics. In Corbett, G., N. M. Fraser, and S. McGlashan (eds) *Heads in grammatical theory.* Cambridge: Cambridge University Press.

Morimoto et al., T. (1993). ATR's speech translation system: ASURA. In *Proceedings of EUROSPEECH'93*, Berlin, Germany.

Peckham, J. and N. M. Fraser (1994). Spoken language dialogue over the telephone. In Nieman, H., R. de Mori, and G. Hanrieder (eds) *Progress and Prospects of Speech Research and Technology.* Sankt Augustin, Germany: Infix. 192–203.

Ripplinger, B. (1994). Concept-based machine translation and interpretation. In *Proceedings of Int. Conf. on Machine Translation, Cranfield, England.*

Wahlster, Wolfgang (1993). Verbmobil: Translation of face-to-face dialogs. In *In Processings of EUROSPEECH'93.* 29–38.