Ontology-Based Dialogue Systems

David Milward & Martin Beveridge
Advanced Computation Laboratory, Cancer Research UK
44 Lincoln's Inn Fields, London, UK
david.milward@linguamatics.com, martin.beveridge@cancer.org.uk

Abstract

This paper examines how far it is possible to replace hand crafted dialogue design with a combination of generic dialogue system components and ontological domain knowledge. Two case studies are presented, one for dialogue-based breast cancer referrals, the other for control of networked home appliances.

1 Introduction

Even modest improvements in the domain knowledge available to dialogue systems should be able to reduce reliance on hand-crafted linguistic components, allowing more flexible dialogues and more easily reconfigurable dialogue systems. This paper provides a broad, if not exhaustive, overview of how domain knowledge can be used by various components of a dialogue system.

Separating out domain knowledge reduces the complexity of the linguistic components. Moreover, if dialogue systems are to be built for domains that have a rich ontological structure (such as the health domain), reuse of existing domain knowledge becomes critical. For cancer applications we are using an existing medical ontology [Ceusters et al., 2001] that contains rich semantic and terminological information. For the home domain, where we did not have access to an existing ontology, we are experimenting with a more minimal hand-crafted ontology.

Section 2 presents some background on the use of ontologies for natural language processing. Section 3 discusses how ontologies can be used to provide flexible dialogue management. Section 4 discusses how terminological information can be used to build language models for speech recognition (with certain provisos), and to assist interpretation and generation. Section 5 describes implementations for two very different domains.

2 Background

There has been considerable effort in recent years in developing large-scale domain ontologies, particularly in the medical and biological sciences. Recent research on description logics [Nardi and Brachman, 2002], has led to a knowledge representation that is weak enough to be highly tractable for rapid search and inference. The work on domain ontologies complements efforts to provide large domain independent lexical ontologies, such as WordNet [Fellbaum, 1998].

WordNet consists of sets of orthographic forms of words with associated sense sets: typically a set of synonyms or "synset" which can be viewed as defining a concept. WordNet also defines a large number of domain-independent lexical relations which can hold between synsets, e.g. hypernym (is-a), part-of, member-of etc. These relations can be linked together to form paths (e.g. hyperonymy chains) between synsets.

Large lexical ontologies such as WordNet have been used in applications such as text summarisation [Chaves, 2001], translation [Knight, 1993] and information extraction [Burke et al., 1995], sometimes directly, sometimes to help build more specialised domain ontologies.

Ontologies have been used within specific dialogue system modules, for example in language generation as the basis of the Systemic Grammar approach [Bateman, 1990]. However, more general use of ontologies is relatively rare. AT&T's "How May I Help You?" system [Abella & Gorin, 1999] uses an inheritance hierarchy of task knowledge covering the available operator services. This is used both to help interpretation, and within dialogue management to generate clarifications. The EMBASSI system [Bűcher et al., 2001] uses a description logic based ontology to evaluate the semantic well-formedness of DRT (Discourse Representation Theory) structures returned by a speech recogniser.

Dahlbäck and Jönsson [Dahlbäck and Jönsson, 1997] have argued that a speech-act based dialogue model could have much wider applicability if combined with sophisticated domain knowledge. They split the domain knowledge into a domain model containing static knowledge e.g. that Heraklion is a town in Crete, and a conceptual model which contains relationships which are more connected with a task such as hotel booking. The conceptual model is used to help interpret user requests such as "I need a shower" as meaning "I need a room with a shower" when uttered in the hotel booking domain (as opposed to e.g. after a football match).

3 Ontologies for Dialogue Management

In dialogue systems that provide relatively open-ended prompts, users may provide responses that are not an exact fit with what is expected. One particular case of this, which has been much examined, is where a user supplies a more informative response involving two separate pieces of information. Consider the following (utterances marked "S" are system utterances and those marked "U" are utterances by the user):

S: Where do you want to leave from?
U: Boston at 8am.

This has been called "mixed initiative" in the VoiceXML literature and is handled in VoiceXML forms by allowing more than one form value to be supplied at once.

However, there are other cases where there is a mismatch between what a user supplies in their response and what was expected by the system. In this section we consider two cases: the first is where the user supplies more information via the use of a single term that is more specific than expected, the second is where the user supplies less information via the use of a more general term. We will also consider how ontological structure can be used to automatically structure a dialogue for maximal coherence.

3.1 Hypernyms and Hyponyms

During dialogue the user may provide a response to a prompt that does not match any of the expected range of responses (as defined by the application domain) but is a hypernym or hyponym of an expected response. In this case the system should be able to discover the relation between the expected terms and the user response in order to resolve the discrepancy.

In the case where the user's reply was a hypernym of an expected term, then their reply can be considered under-specified and the system can issue clarificatory questions in order to obtain a more detailed reply. Consider the following example from the cancer domain:

S: Do you have a family history of chronic disease?
[*system expects: lung cancer, leukemia, sarcoma, …*]
U: Yes, <u>cancer</u>
S: What type of cancer?
U: <u>Lung cancer</u>.

In this example the system expected a specific disease but the user replied with a more generic term. The system therefore formulated a more specific question in order to elicit an answer at the expected level of the is-a hierarchy. Hypernyms are similarly treated in a generic fashion in the AT&T system via the use of a disambiguation operation.

In the case where the user's reply was a hyponym of an expected term then their reply can be considered over-specified and the system can (a) find a more general related term which matches the expected responses in order to answer the current question and (b) avoid asking subsequent more specific questions that have already been answered. For example

S: Have you had any chronic diseases?
[*system expects: cancer, hypertension, diabetes, …*]
U: Yes, <u>Leukemia</u>
S: # What kind of cancer?

In this example the system expected a more generic answer to the question (such as "cancer"), but the user replied with a more specific term. The system, however, was able to match the answer to the question on the basis that leukemia is a cancer, which is a chronic disease. The system avoids asking what kind of cancer since this has already been answered and therefore would be pragmatically ill-formed (indicated by a "#"). Instead the system continues with the next item in its plan.

3.2 Part-Whole Relationships

During dialogue the user may provide a response to a prompt that does not match any of the expected range of responses (as defined by the application domain) and furthermore is not a hypernym or hyponym of an expected response, but is associated with one or more of the expected responses by a non-is-a link which imposes an ordering on concepts (e.g. part-whole links). In this case the system should be able to discover the associative relation between the expected terms and determine whether the supplied term is more or less specific than the expected terms according to the ordering imposed by the relation.

In the case where the user's reply can be considered under-specified with respect to the expected terms the system can issue clarificatory questions in order to obtain a more specific reply. For example:

S: Where does it hurt?
[*system expects: elbow, wrist, shoulder,…*]
U: In my <u>arm</u>.
S: Where in your arm?
U: In my <u>elbow</u>.

In this example the system initially expected a more specific body-part than the user supplied, but recognized that the supplied term "arm" was related to the expected terms in a part-whole hierarchy and was more general then the expected terms. The system therefore did not simply repeat the original question "where does it hurt?" but instead formulated a new question to elicit an answer at the expected level of the part-whole hierarchy.

In the case where the user's reply can be considered over-specified with respect to the expected terms the system can (a) find a more general term which matches the expected responses in order to answer the current question and (b) avoid asking subsequent more specific questions that have already been answered. For example:

S: Where does it hurt?
[*system expects: arm, leg, head, …*]
U: In my elbow
S: # Does your arm hurt?

3.3 Dialogue Coherence

In system-initiated dialogue, questions are grouped together at design time. Even in systems allowing some mixed initiative, this rarely affects the ordering of subsequent questions. Consider, the following dialogue:

S: What is the patient's sex?
U: Female with severe nipple discharge
S: What is the patient's age?
U: Fifty five
S: Is the discharge bilateral?
U: No

Here we assume that the system would normally ask for gender first, followed by age. It is intelligent enough not to ask whether or not the patient has nipple discharge, but otherwise the ordering of the questions is kept fixed. We can achieve much more natural dialogues if we cluster questions dynamically following changes by the user. Consider the following:

S: What is the patient's sex?
U: Female with severe nipple discharge
S: Is the discharge bilateral?
U: No
S: What is the patient's age?
U: Fifty five

This appears to be a much more natural exchange, with the system immediately asking the follow-on question concerning nipple discharge. How can this be achieved? Again, ontological information plays a part. Questions which elaborate a previous question, either by asking about a particular attribute, or by asking for more specific information (determined by the is-a or part-whole hierarchy) will ideally appear straight afterwards, similar to elaboration in well structured texts [Mann and Thompson, 1988].

4 Ontology Use in Recognition, Interpretation and Generation

4.1 Recognition

Ontologies provide terminological information useful for generating context specific (dynamic) language models. For example, consider the following question:

S: Does the patient have cardiac arrhythmia?
[*system expects: atrial fibrillation, sinus tachycardia, AV block, …*]

Here, we can use synonym sets to create a language model containing not only "atrial fibrillation" but also all its synonyms e.g. "auricular fibrillation". We have used this both for grammar based language modeling, and for populating a class in a class-based statistical language model.

There are some limitations to this approach, however. Terminological ontologies tend to provide variations in the way complex terms are expressed e.g. including both "nipple retraction" and "retraction of the nipple". However, a grammar based on a disjunction of these two terms will not recognize "retraction of the left nipple" or "the nipple has retracted".

To achieve a more comprehensive recognition grammar we need to use a more semantically based approach to term recognition. This is the approach we use in the home domain. The ontology associates terms with kinds of devices and rooms, but not with individual devices. For example, there is a term for "light" and a term for "kitchen" but no term for "kitchen light". The recogniser can cope with instantiations of e.g. "<room> <device-type>", "<device-type> in the <room>" and "<device-type> in the back <room>".

4.2 Interpretation

In order to interpret user responses it is necessary to map lexical terms into domain concepts. Furthermore, one term may map to several concepts (polysemy). For example, consider the following exchange:

S: Are there any skin changes?
U: There is some distortion

The term "distortion" might refer to various concepts in the ontology (e.g. skin distortion, nipple distortion etc) and so the system has to determine which concept is intended. Since the context is skin changes, the ontology can be searched to find a concept which is related to skin change and for which "distortion" is a term.

4.3 Generation

In generating a term for a particular concept, a terminological ontology will provide one or more terms. There may be reasons for choosing one synonym rather than another (for example, to follow the usage of a user). If not, many ontologies also include a notion of 'preferred term' which can be used by default.

The upper ontology may also provide appropriate concepts for sentence generation in the same way as the Upper Model [Bateman, 1990]. This will be described later for the medical domain.

5 Case Studies

These ideas have been evaluated by looking at two disparate domains. The first is a medical decision support application. The second is home control.

5.1 Decision Support for Cancer Referrals

The cost of managing patients with chronic diseases is progressively increasing. In order to improve the capability of healthcare institutions to manage chronic diseases, it has been recognized that medical guidelines must be implemented more effectively, which will often require patients' conditions to be monitored and assessed more frequently. A cost-effective way of achieving this may be to use communication technologies such as the Internet and telephone to allow patients to provide data (e.g. blood pressure measurements for hypertension patients) and to receive advice (e.g. revising insulin levels for patients with diabetes). This could also offer patients more control of their own care, by allowing them to obtain advice whenever and as often as they need it. Voice interaction seems a particularly suitable approach as speech is a natural way for people to communicate and does not require the patient to use any technology other than a telephone. Hence, the technological barrier is reduced and the potential reach of the system is increased (especially among the elderly).

Natural language dialogue systems therefore seem to have great potential for assisting in delivery of healthcare services, but the state of the art in medical applications of dialogue systems is extremely limited. One reason for this may be that medicine is a more challenging domain for implementing dialogue systems than many of those that have been investigated so far, e.g. route planning, flight booking etc which generally involve simple information look-up. In building a patient home-monitoring system, however, the range of dialogue will typically be much more varied and the clinical service must make use of expert knowledge and be able to perform complex reasoning and decision-making in responding to the user. Overall, therefore, it seems necessary to better integrate technologies such as medical guidelines and advice systems with the dialogue system so that the dialogue can be closely related to the underlying clinical context and goals.

Cancer Research UK (CR-UK) has developed a dialogue system as part of the HOMEY project (Home Monitoring through an Intelligent Dialogue System, EC project IST-2001-32434) that supports electronic referrals for breast cancer [Beveridge and Milward, 2003a; Beveridge and Milward, 2003b]. The system is intended to be used by medical General Practitioners (GPs) to determine whether a patient with suspected breast cancer should be referred to a cancer specialist or treated within-practice. The dialogue system therefore needs to be closely integrated with technologies for representing and enacting clinical guidelines, in this case the PRO*forma* process specification system [Fox and Das, 2000], and with medical domain ontologies, in this case an ontology provided by Language & Computing n.v. [Ceusters et al., 2001].

Architecture

In order to allow integration with the clinical domain representations, the dialogue model is divided into high- and low-level representations. The low-level representation defines a finite-state network of communication acts and is represented by a VoiceXML specification. The high-level representation captures information regarding the intentional and informational structures underlying the dialogue, along with its current attentional state [Grosz and Sidner, 1986; Hobbs, 1996; Moore, 1995]. The information in the high-level representation is in turn derived from the underlying domain specification, with the *intentional* structure deriving from tasks in the medical guideline and the *informational* structure deriving from the medical ontology. In order to make use of these representations in a practical system we have employed a multi-level architecture (similar to 3-layer hybrid agent architectures [Gat, 1998]) as shown in Figure 1.

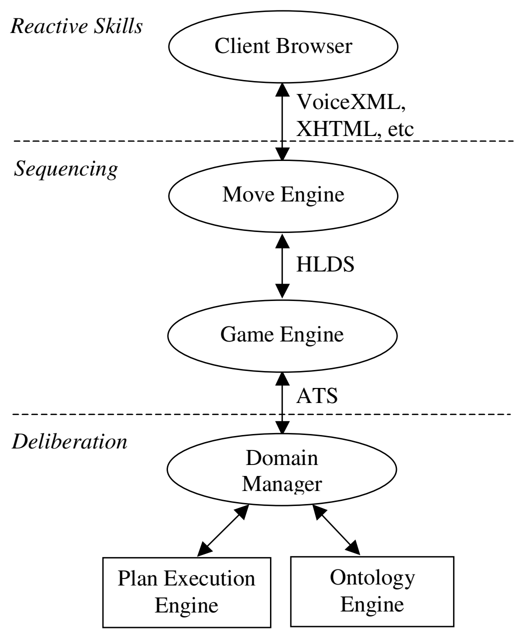

Figure 1: The CR-UK Dialogue System Architecture.

The deliberative layer is provided by a domain manager which creates an abstract task specification (ATS) based on the outputs of the plan execution and ontology engines. The sequencing layer includes a Game Engine which determines the conversational games [Kowtko and Isard, 1993; Lewin, 2000] to be played in order to complete the tasks in the ATS, and orders them for maximal dialogue coherence. The resulting High-Level Dialogue Specification (HLDS) is used by a Move Engine to generate the sequence of low-level communicative acts (moves) that can be made by either participant at the current point in the dialogue. The reactive layer interprets the low-level specification (VoiceXML) in order to complete the specified dialogue segment, and handles low-level reactive behavior required to support communication, e.g. repeating prompts, changing the speech volume etc.

The following sections focus on the role of ontological knowledge in this architecture, and the way in which the ideas presented in Sections 3 & 4 are employed in the CR-UK dialogue system.

Hypernyms and Hyponyms

As described in Section 3, the CR-UK system uses ontological subsumption (is-a) relations to reason about under- and over-specified user replies, such as the use of hypernyms or hyponyms of expected terms. For example, in the following interaction the user provides an over-specified reply that refers to a hyponym of the term presented in the question:

S: Does the patient have a breast nodule?
U: She has a <u>breast induration.</u>
S: Ok.

The concept related to the term "breast induration" is subsumed by the concept related to the term "breast nodule" (i.e. "breast induration" is a hyponym of "breast nodule") and so the system can infer from the ontology that the patient does have a breast nodule. The user's reply can therefore be successfully accommodated as an answer to the system's question. Conversely, the following interaction demonstrates an under-specified response by the user:

S: Does the patient have a lesion?
[system expects: skin ulcer, gastric ulcer, etc …]
U: She has an <u>ulcer</u>.
S: What kind of ulcer does the patient have?
U: A <u>skin ulcer</u>

In this case the concept related to the term "ulcer" subsumes all the concepts that the system can accept as valid responses and so the system attempts to clarify what exactly was meant. The subsequent reply "skin ulcer" is related to one of the concepts the system expected, and can therefore be accommodated as an answer to the original question.

Part-Whole Relationships

Part-whole relations could be used in a similar way to hyperonymy relations to expand the range of replies that can be accommodated by the system. However, as such relations do not frequently occur in the breast cancer referrals domain they are not currently used.

Dialogue Coherence

In the CR-UK system the overall dialogue structure is only *partially* determined by the intentional structure (derived from the task structure of the underlying medical guideline). Ontological knowledge is also used in order to ensure that the dialogue is coherent with respect to the relations between the topics under discussion. These then give rise to rhetorical relations between segments in the rendered dialogue [Beveridge and Milward, 2003a; Stent, 2000]. For example, the intentional structure of a dialogue may define intentions and relations such as those below (where *DOM* indicates a dominance relation [Grosz and Sidner, 1986]):

I_1 = Intend(S, Intend(H, Know(S, **clinical details**)))
I_2 = Intend(S, Intend(H, Know-if(S, **nipple discharge**)))
I_3 = Intend(S, Intend(H, Know-if(S, **bilateral nipple discharge**)))
I_1 *DOM* I_2
I_1 *DOM* I_3

From this we can infer that I_2 and I_3 must both be satisfied in order for I_1 to be satisfied but the intentional structure only provides a *partial* ordering - it does not specify which of I_2 and I_3 should receive focus first in the attentional state. Suppose however that the domain ontology contains a specification that **bilateral nipple discharge** is subsumed by **nipple discharge**. In this case we can infer the following information relation: I_3 *elaborate* I_2 [Mann and Thompson, 1988]. Hence I_2 is the nucleus of an elaboration relation of which I_3 is the satellite, and hence I_2 should receive focus first. This relation in the information structure (derived from the medical ontology) therefore allows a total ordering of dialogue intentions to be constructed in the attentional state. Furthermore, if at any point in the dialogue the user addresses I_2 (for example by making use of mixed-initiative) then the dialogue system should consider I_3 as a priority for receiving focus next (assuming that it does not violate constraints imposed by the intentional structure). This allows the system to shift to the topic introduced by the user and pursue it immediately, hence preserving dialogue coherence.

Recognition

The terminological information in the medical ontology is used to generate a grammar-based language model with rules defining synonym sets for individual concepts. These are then combined with general rules that provide a template into which synonym sets can be slotted. E.g. consider the case where the dialogue system has intentions such as the following:

I_1 = Intend(S, Intend(H, Know-if(S, **breast cyst**)))
I_2 = Intend(S, Intend(H, Know-if(S, **skin ulcer**)))

The language model in this case should contain at least the following synonym sets:

breast cyst: {"breast cyst", "cyst of breast", "simple breast cyst", "simple cyst of breast", …}

skin ulcer: {“skin ulcer”, “ulcer of skin”, “skin ulceration”, “ulceration of skin”, “cutaneous ulcer”, …}

In order to allow the accommodation of hypernyms and hyponyms (as described earlier), synonym sets for concepts that subsume or are subsumed by these concepts are also retrieved from the ontology. E.g. in the case of **breast cyst**, the ontology might contain the following relations: **recurrent breast cyst** is-a **breast cyst**, **lactocele** is-a **breast cyst**, **breast cyst** is-a **cyst, breast cyst** is-a **breast mass**. The following additional synonym sets should therefore also be included in the language model:

recurrent breast cyst: {“recurrent breast cyst”, “recurrent cyst of breast”, … }
lactocele: {“lactocele”, “lacteal cyst”, “milk cyst”, …}
cyst: {“cyst”, “cysts”, “simple cyst”, …}
breast mass: {“breast mass”, “mass in breast”, … }

A simplified version of the template rule for recognizing a fully formed utterance is of the following form:

<*reply*> = (“he” | “she”) “has” (“a” | “an”) (<*breast cyst*> | <*recurrent breast cyst*> | <*lactocele*> | <*cyst*> | <*breast mass*> | <*skin ulcer*> | …)

where <*cyst*>, <*breast cyst*> etc are grammar rules that accept all the terms in the synonym sets for the associated concepts (**cyst**, **breast cyst** etc.).

Interpretation

Given the above language model, the interpretation process simply involves matching terms to domain concepts as specified in the ontology. Those concepts can then be matched against the concepts expected by the system and ontological relations calculated in order to resolve under- or over-specified replies etc. (as described earlier).

Generation

Currently we also make use of upper ontology concepts as categories for text generation. These include: ***property*** (e.g. **age**, **sex**), ***material entity*** (e.g. **cyst**, **ulcer**) and ***material process*** (e.g. **distortion**, **retraction**). These allow the text generator to choose appropriate templates for generation, based on which category subsumes the topic of the utterance to be rendered. E.g. the templates for generating requests are of the following form:

property: “What is the patient’s” **age**, **sex**…?
material entity: “Does the patient have a” **cyst**, **ulcer**…?
material process: “Does the patient have any” **distortion**, **retraction**…?

The term used to represent a concept in sentence generation is determined simply by defining one of the terms in the ontology as a “preferred term” as described in Section 4.3.

Example Dialogue

An example of a full dialogue with the system, illustrating some of the concepts described above, is given below. Note that the generation and interpretation components developed so far are relatively simple and currently lack mechanisms (such as use of anaphora) that would allow for a better-articulated conversation. These components will be improved in later versions of the system.

S: Hello, this is the Cancer Research UK cancer referrals advice system…
U: Hi
S: What is the patient's age?
U: She is thirty five, female, with nipple retraction, a breast abscess, and an ulcer.

[The user answers the system’s question, but also provides additional information (underlined). This is mostly accommodated by matching it to questions the system was planning to ask. The concept ***ulcer****, however, is too general to be directly matched to any pending question and so the system attempts to clarify what exactly was meant…]*

S: What kind of ulcer does the patient have?

[The system expects ***skin ulcer****,* ***gastric ulcer****,* ***gastro-intestinal ulcer*** *etc.]*

U: A cutaneous ulcer

[The system can accept the term “cutaneous ulcer” as it is in the synonym set for ***skin ulcer****. The clarification side-sequence is therefore completed and the system can return to its original dialogue plan…]*

S: Does the patient have an acquired deformity of nipple?
U: No, but she has some nipple discharge

[The user again provides additional information (underlined) which is matched against pending questions. In this case the system also has several pending questions regarding more specific concepts that are subsumed by ***nipple discharge****. It therefore re-plans the dialogue in order to pursue the topic introduced by the user and obtain the more specific information required…]*

S: Does the patient have bilateral nipple discharge?
U: No, she has bloodstained nipple discharge
S: Does the patient have cloth-staining discharge?
U: No

[The system has obtained all the required information related to ***nipple discharge*** *and so introduces a new topic…]*

S: Does the patient have a breast nodule?

U: She has a breast induration and she has a cyst

*[The concept **breast induration** is-a **breast nodule** so the system can infer that the patient does have a breast nodule, and hence the reply can be accommodated. The concept **cyst**, however, is too general to be directly accommodated and so the system attempts to clarify this reply…]*

S: What kind of cyst does the patient have?
U: What do you mean?

[The user in turn requests clarification. In this case, the system explains what types of cyst are relevant…]

S: I mean: do they have a breast cyst, retention cyst or sebaceous cyst?
U: She has a breast cyst

[The user provides a reply to the original question, hence implicitly acknowledging the system's explanation and ending the clarification side-sequence.]

S: Does the patient have any intractable pain?
U: No
S: Does the patient have an asymmetrical breast nodularity?
U: Yes, and she has a breast lump

[The user provides additional information (underlined) which matches a pending question and so can be directly accommodated.]

S: Does the patient have nipple eczema?
U: No

[The system makes a decision, based on standard guidelines, as to whether the patient should be referred to a specialist or treated in-practice. The result of the decision-making process is provided as advice to the user.]

S: This patient is a woman with possible breast cancer who should be referred to a specialist.
U: Right, thanks…

[The user has acknowledged the system's advice and so the dialogue is complete.]

S: Ok. Thankyou for calling. Goodbye.
U: Bye.

5.2 Home Device Querying and Control

Spoken dialogue provides a natural way of communicating with individual or networked home devices. This has immediate applicability for the elderly or physically disabled. As homes become networked as a matter of course, and we become used to services being provided by multiple devices, spoken dialogue can also provide a good alternative to universal remote controllers. For example, there may be no one box comprising the TV, but several devices e.g. a flat screen, speakers, receiver etc. It then becomes natural to think in terms of services rather than devices e.g. "show me the news", "give me mood lighting", or "cool the room". We may also want to program across devices e.g. "turn on the hall light when the door is opened".

The Linguamatics dialogue system was inspired by work on the EU project D'Homme. The D'Homme project [DHomme, 2001] investigated spoken dialogue in the home machine environment, and was particularly concerned with how to deal with plug and play i.e. how to allow for new devices to be added to a system without having to reprogram the dialogue manager. The Linguamatics dialogue system goes one step further. All domain knowledge is isolated within the domain ontology, which can be reconfigured dynamically. The dialogue manager is fully generic, dealing with (partial) commands or queries, and asynchronous tasks introduced by devices (e.g. by sensor activation). Reconfiguration is achieved by changing the ontology, the language model and the generation templates.

The ontology used for home control includes lexical information and synonyms for individual concepts such as **light**, **kitchen** etc. There are two main hierarchical relationships, is-a and in. The is-a relationship contains general information e.g. **light** is-a **device** and instance (A-box) information about the class of individual devices. The in hierarchy describes which devices are in which rooms, which rooms are upstairs etc. Unlike a terminology inspired ontology, concepts are not included for complex terms unless absolutely necessary. For example, an item such as "the back bedroom light" is treated as an instance of a light, having the location "back bedroom", without introducing a new concept "back bedroom light" into the ontology.

The system allows users to query and control multiple networked appliances using relatively free-form speech. Dialogues are usually initiated by the user asking to perform commands e.g. "turn off all the lights" or make queries e.g. "have I left the cooker on?". The system takes over for clarification, help, or to establish extra parameters for a command. Dialogues may also be initiated by devices. For example, a sensor may be triggered, causing the current dialogue to be interrupted.

The following sections focus on the way in which the ideas in Sections 3 & 4 are employed in the Linguamatics dialogue system.

Ontologies for Dialogue Management

For the cancer application, the most common mismatches between user responses and system expectations are due to the guideline enactment engine requiring information at a particular level of granularity (e.g. cancer vs. leukemia or vice-versa). In the home domain, the level of granularity tends to be much clearer (e.g. a particular device in a state of **on** or **off**). Where hyponymy does occur in the home domain is in examples such as the following:

S: Which service do you require?
U: Air conditioning

Air conditioning is not a top-level service, but is a hyponym of the top-level service, temperature control. This example corresponds to skipping one or more levels of menu structure. Although menu structures and ontological structure are not forced to correspond one-to-one, we do assume that a daughter of a menu item is subsumed, directly or indirectly, by its parent.

Hyperonymy (under-specification) is most commonly seen in reference resolution e.g.

S: Which device?
U: The tv
S: Do you mean the living room tv or bedroom tv?
U: The living room tv

The system correctly determines that **tv** is a hyponym of **device** according to the is-a hierarchy for devices, so is an adequate reply, but not as fully specified as the system requires. It then looks to see if there is a single instance in the ontology for this class. If not, as in this case, it asks for a further clarification of the particular television intended.

The system does not require parameters to commands to be provided by the user in a particular order, and allows for multiple parameters to be specified at once. However, it uses a default order, and does not perform reordering according to the ontological structure as described in Section 3.3. We would expect to incorporate this, in future versions of the system.

Recognition

Language modeling uses both grammar-based and n-gram techniques. Evaluation in the D'Homme project for recognition of user initiated queries and commands suggested that statistical n-gram language models give better results for untrained users, and are better at picking up partial information than grammar based approaches [Knight et al. 2001]. The Linguamatics system therefore uses n-gram language modeling for user initiated queries and commands, with grammar based approaches used for more constrained contexts, e.g. to interpret answers to clarification questions, such as "do you mean the front door or the back door?".

The class-based statistical language model was derived by collecting a corpus of user utterances and defining a set of classes. For example, the class "device-type" is populated by the term corresponding to "device" and all the terms corresponding to the sub-concepts of "device" in the ontology e.g. "cooker", "light" etc. Corpus examples involving individual class terms, such as "turn on the light", are generalized to "turn on the <device-type>" before creating the statistical language model. Classes are redefined dynamically (where necessary) by finding all the terms in the ontology corresponding to instances of the class.

Interpretation

The system uses the synonyms provided in the ontology to provide candidate concepts for the interpreter. Interpretation proceeds by a process of semantic-based composition where relationships in the ontology are used to propose appropriate combinations of terms cf. [Milward, 2002]. For example, a term such as "back bedroom light" is interpreted as [[light] *in* [bedroom *orientation* back]].

Generation

Sentence generation is template based. Generation of referring expressions uses the ontology. An entity is described by its most specific class e.g. "light" and distinguished from other items of that class according to its properties, such as its location. A term successfully describes the item if there are no other candidate items with this description c.f. [Dale and Haddock, 1991].

5.3 Evaluation

We have considered the use of ontologies to provide dialogue systems in two very different domains, in one case reusing an existing ontology, in the other building an ontology from scratch.

Our aim was to provide more easily reconfigurable dialogue systems by describing the structure of the domain and the tasks. This is in contrast to finite state, or VoiceXML style dialogues where the emphasis is on describing the possible dialogue interactions.

In the case where there is no existing ontology, a key question is whether it is easier to build a new ontology, rather than building a description of all possible dialogue interactions. Our experience from the home control application is that ontology building for a particular application can be relatively easy. Moreover, you need an ontology or some similar knowledge representation component anyway if you are to have reference resolution capability.

In the case where there is an existing domain ontology, a key question is whether it is likely to be suitable for use by a dialogue system, or whether adaptation of the ontology will be more work than starting from scratch. Our experience in the cancer domain has been very positive, with a good fit between the particular task and the information needed by the dialogue manager from the ontology.

In both dialogue systems hand crafting is kept at a minimum. In the Cancer Research system the language model and the generation templates are specific to patient care applications. In the Linguamatics system, the ontology was hand crafted, and the generation templates are specific to each task. The language model is derived statistically.

Both dialogue systems have been recently developed, and we have not yet undertaken any formal evaluation. The home control application will be field tested in 50 homes in the second half of 2003. We expect the cancer referrals application to be evaluated by OpenClinical users (http://www.openclinical.org/). Evaluation of this data will provide a better idea of how important flexible dialogue phenomena (such as the use of hypernyms or hyponyms) are for ordinary users.

6 Conclusions

In this paper we have shown how domain and terminological knowledge can be used by linguistic components ranging from recognition to dialogue management. This allows for more general linguistic components, and hence more easily reconfigurable systems. We see this as an important step in moving from the scripted dialogues deployed in most current commercial systems, to a new generation of practical systems based on domain knowledge and task descriptions.

Acknowledgments

The work on medical systems described in this paper was funded by the European Union under the 5th Framework Project Homey. The work on home control has had support from the EU 5th Framework Project, Siridus, and the UK Department of Trade and Industry Next Wave Programme. We would like to thank John Fox at Cancer Research and our partners in Homey for many useful discussions.

References

[Abella and Gorin, 1999] A. Abella and A. Gorin. Construct Algebra: Analytical Dialog Management. *Proceedings of the ACL*, Washington D.C., June 1999.

[Bateman, 1990] J. A. Bateman. Upper Modeling: A General Organization of Knowledge for Natural Language Processing. *Proceedings of the Workshop on Standards for Knowledge Representation Systems*, Santa Barbara, March 1990.

[Beveridge and Milward, 2003a] M. A. Beveridge and D. Milward. Definition of the High-Level Task Specification Language. *Deliverable D11, EU 5th Framework Programme, HOMEY Project*, IST-2001-32434, March 2003.

[Beveridge and Milward, 2003b] M. A. Beveridge and D. Milward. Combining Task Descriptions and Ontological Knowledge for Adaptive Dialogue. To appear in *Proceedings of the 6th International Conference on Text, Speech and Dialogue (TSD)*, Ceske Budejovice, Czech Republic, September 2003.

[Bücher et al., 2001] K. Bücher, Y. Forkl, G. Gorz, M. Klarner, and B. Ludwig. Discourse and application Modeling for Dialogue Systems. *Proceedings of the KL-2001 Workshop on Applications of Description Logics*, Vienna, Austria, September 18th, 2001.

[Burke et al., 1995] R. D. Burke, K. J. Hammond, and J. Kozlovsky. Knowledge-based information retrieval from semi-structured text. *AI Applications in Knowledge Navigation and Retrieval. Papers from the 1995 AAAI Fall Symposium (Tech. Report FS-95-03)*, AAAI Press, Menlo Park, pp. 15 – 19, 1995.

[Ceusters et al., 2001] W. Ceusters, P. Martens, C. Dhaen, and B. Terzic. LinkFactory: an Advanced Formal Ontology Management System. *Proceedings of the Interactive Tools for Knowledge Capture Workshop, KCAP-2001*, Victoria B.C., Canada, 2001.

[Chaves, 2001] R. P. Chaves. WordNet and Automated Text Summarization. *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium, NLPRS*, Tokyo, Japan, 2001.

[Dale and Haddock, 1991], R. Dale & N. Haddock. Content determination in the generation of referring expressions. *Computational Intelligence* 7(4), 1991.

[Dahlbäck and Jönsson, 1997] N. Dahlbäck and A. Jönsson. Integrating Domain Specific Focusing in Dialogue Models. *Proceedings of EuroSpeech '97*, Rhodes, Greece, 1997.

[DHomme, 2001] Dialogues in the Home Machine Environment. Technical Showcase and Deliverables. Available from http://www.ling.gu.se/projekt/dhomme.

[Fellbaum, 1998] C. Fellbaum (Ed.). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA, 1998.

[Fox and Das, 2000] J. Fox, and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*, AAAI Press, Menlo Park, CA, and MIT Press, Cambridge, Mass, 2000.

[Gat, 1998] E. Gat. On Three-Layer Architectures. In Artificial Intelligence and Mobile Robots, D. Kortenkamp, R. Bonasso and R. Murphy (eds), AAAI Press, Menlo Park, CA, 1998.

[Grosz and Sidner, 1986] B. Grosz, and C. Sidner. Attention, Intention and the Structure of Discourse. *Computational Linguistics* 12(3):175-204, 1986.

[Hobbs, 1996] J. R. Hobbs. On the Relation between the Intentional and Informational Perspectives on Discourse. In *Burning Issues in Discourse: an Interdisciplinary Account*, E. Hovy and D. Scott (eds), Springer-Verlag, Berlin, 1996.

[Knight, 1993] K. Knight. Building a large ontology for machine translation. *Proceedings of the ARPA Human Language Technology Workshop*, Princeton, 1993.

[Knight, 2001] S. Knight, G. Gorrell, M. Rayner, D. Milward, R. Koeling, I. Lewin. Comparing grammar-based and robust approaches to speech understanding: a case study. *Proceedings of Eurospeech 2001, 7th European Conference on Speech Communication and Technology,* Aalborg, Denmark 3-7 September 2001.

[Kowtko and Isard, 1993] J. Kowtko and S. Isard. *Conversational Games Within Dialogue*, Research Paper 31, Human Communication Research Centre, Edinburgh, 1993.

[Lewin, 2000] I. Lewin. A Formal Model of Conversational Game Theory. *Proceedings of Gotalog '00, 4th Workshop on the Semantics and Pragmatics of Dialogues*. Gothenburg, 2000.

[Mann and Thomson, 1988] W. D. Mann, and S. A. Thompson. Rhetorical Structure Theory: Towards a functional theory of text organization. *Text* 8(3):243-281, 1988.

[Milward, 2002] Exploiting the Advantages of Task and Linguistically Orientated Dialogue Management. *Siridus Deliverable D4.4*. October 2002. Available from http://www.ling.gu.se/projekt/siridus.

[Moore, 1995] J. Moore. The Role of Plans in Discourse Generation, In *Discourse: Linguistic, Computational, and Philosophical Perspectives*, D. Everett and S. G. Thomason (Eds.), 1995.

[Nardi and Brachman, 2002] D. Nardi, and R. J. Brachman. An Introduction to Description Logics. In *The Description Logic Handbook*, F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider (eds), Cambridge University Press, pp. 5-44, 2002.

[Stent, 2000] A. Stent. Rhetorical Structure in Dialog. *Proceedings of the 2nd International Natural Language Generation Conference (INLG),* 2000.