Some empirical findings on dialogue management and domain ontologies in dialogue systems – Implications from an evaluation of BIRDQUEST

Annika Flycht-Eriksson
Department of Computer and
Information Science
Linköping University, Sweden
annfl@ida.liu.se

Arne Jönsson
Department of Computer and
Information Science
Linköping University, Sweden
arnjo@ida.liu.se

Abstract

In this paper we present implications for development of dialogue systems, based on an evaluation of the system BIRDQUEST which combine dialogue interaction with information extraction. A number of issues detected during the evaluation concerning primarily dialogue management, and domain knowledge representation and use are presented and discussed.

1 Introduction

In the field of Question Answering (Q&A), Information extraction (IE) techniques have been used successfully when it comes to handling simple factoid questions, but the Q&A approach has yet not reached the level of sophistication for handling connected dialogue as is present in dialogue systems tailored to background systems with structured data. Dialogue capabilities allow for more precise formulation of information requests and more natural interaction. The challenge is to combine the IE techniques and some of the features of Q& approaches with dialogue systems (Burger et al., 2001). By a successful combination of these techniques, users would be allowed to access information derived from a large set of, initially unstructured, documents, using dialogue functionalities, such as a dialogue history and clarification requests.

We have developed a first version of such a combined system, BIRDQUEST (Jönsson and Merkel, 2003), which supports dialogue interaction to access textual data in a bird encyclopaedia. The source data is initially provided as unstructured text but refined with IE techniques to be used within a dialogue system framework. As a basis for many of the tasks in the system domain knowledge represented in an ontology is utilised.

To assess the approach and get insights into what areas need further improvement an evaluation of the system has been carried out. In this paper the results of this evaluation are presented together with a discussion of implications for development of dialogue systems with focus on dialogue management and the use of domain ontologies.

2 Combining IE with dialogue interaction in a system

Combining dialogue interaction with information extraction has several benefits; dialogue is a natural and efficient means of interaction and with IE techniques information can be retrieved from unstructured information sources that are otherwise hard to manage and search for a user. A possible way of merging these two in a practical system is to have two components, an information processing component and an interaction component that, as a basis for their tasks, use a set of shared knowledge sources that define the scope of the language and domain.

2.1 The Information Processing Component

The Information Processing Component takes collections of unstructured or semistructured documents and transforms them into structured information that can be used by the Interaction Component in the interaction with the user. The transformation utilise IE techniques, and the documents are analysed in several ways going through lexical and morphological, syntactical, and semantical analysis (Sullivan, 2001).

A wide variety of pattern extraction rules are used to identify the relevant information as slots and fillers. The objective is to fill the database with relevant information and ignore text segments that do not meet the needs of the users. Figure 1 illustrates how unstructured text is transformed into slot and filler type information in the database.

Original text

```
Black-throated diver
Gavia arctica
58-73 cm, wingspan 110-130 cm.
In breeding plumage the head is gray
and the throat is black, the sides
of the throat striped in black and
white. [...]
```

Extracted information

NAME:	Black-throated diver
LATIN_NAME:	Gavia arctica
MAX_WING:	130
MIN_WING:	110
MAX_HEIGHT:	73
MIN_HEIGHT:	58
BR_PLUMAGE:	”the head is gray and the throat is black, the sides of the throat striped in black and white.”

Figure 1: Original text passage from the text book and the corresponding entry in the database (translated from Swedish).

2.2 The Interaction Component

The Interaction Component is responsible for the dialogue with the user. It collaborates with the user to produce a query and access the structured information sources to retrieve an answer to the query. The interaction component in BIRDQUEST is based on the MALIN framework (Dahlbäck et al., 1999). MALIN is a modularised dialogue system and it separates dialogue management (DM) from domain knowledge management (DKM) (Flycht-Eriksson and Jönsson, 2000). The former handles the dialogue whereas the latter handles access to various background information sources.

The Dialogue Manager is responsible for controlling the flow of the dialogue by deciding how the system should respond to a user utterance. This is done by inspecting and contextually specifying the information structure produced by an interpretation module. The MALIN dialogue model classifies the discourse segments by general speech act categories, such as *question* (Q) and *answer* (A), rather than specialised (cf. (Hagen, 1999)), or domain related (Alexandersson and Reithinger, 1995). The dialogue manager instead utilise the focal parameters to control interaction (cf. (Jokinen et al., 1998; Denecke, 1997; Jönsson, 1995)). In MALIN dialogue history is represented in dialogue objects with a parameter termed Objects, which identify a set of primary referents, and the parameter Properties which denote a complex predicate ascribed to this set. In BIRDQUEST Objects are normally birds and Properties model information about the birds, such as appearance, number of eggs and feed.

The Domain knowledge manager receives requests from the dialogue manager and process them further using domain knowledge, for example, disambiguation and mapping of vague concepts to ones more suitable for database access. It then retrieves and coordinates information from available information sources, such as data and knowledge bases. If a request is under-specified or contains inconsistencies from the domain knowledge manager's point of view, a specification of what clarifying information is needed will be returned to the dialogue manager to help the formulation of a clarification question to the user.

2.3 Knowledge sources

As a basis for the processing of documents and user queries a number of knowledge sources are utilised. Some are highly specialised and only used by one or a few submodules of a component, for example the dialogue model in the Interaction Component, while others are more general and used for several tasks in both components. These shared knowledge sources comprise lexicon, grammar, and domain ontologies. Building lexicon and grammars to be used for different tasks also involves several challenges but will not be further discussed in this paper.

The term ontology is used very differently in various areas of computer science, ranging from simple taxonomies, meta data schemes, to logical theories. A general and commonly used definition given by Gruber (1993) is that *"An ontology is a formal, explicit specification of a shared conceptualisation"*. A more practical view is to consider an ontology as *"a world model used as a computational resource for solving a particular set of problems"* (Mahesh and Nirenburg, 1995), i.e. a database with information about what categories (or concepts) exist in the world/domain, what properties they have, and how they are related to one another.

An ontology provides a common vocabulary that can be used to state facts and formulate questions about the domain. Constructing an ontology that can be shared by the Information Processing Component and the Interaction Component then gives us a possible way to bridge users' expression and queries to the information contained in the unstructured documents.

3 Constructing the domain ontology

A challenge when constructing a shared domain ontology lies in capturing and including two different conceptualisations of the domain, the one present in the information sources and the one users have. The shared ontology for the BIRDQUEST system was developed based on the analysis of two different types of empirical material, a bird encyclopaedia and a question corpus. The corpus consists of more than 250 questions about birds. It was collected by The Swedish Public Service Television Company on a web site for one of their nature programs, where the public could send in questions, i.e. it is not a dialogue corpus.

The analysis of the empirical material focused on identifying objects and properties, which in turn were organised using hyponym relations. From the encyclopaedia a conceptualisation underlying the structure and presentation of information that were to be extracted by the Information Processing Component was constructed. The result was a system-oriented domain ontology representing experts' view of the domain. The question corpus yielded a user-oriented conceptualisation of the domain, thus providing a non-expert view of the domain useful for the interaction component. These two conceptualisations were then merged to form a shared domain ontology for all components of the system.

The users' view of the domain as reflected in the questions seemed to correspond to the one found in the reference book, most objects and properties were the same, but there were two aspects that deviated. The first concerned the classification of birds and the second the granularity of the properties of birds.

- Users sometimes utilised another way of categorising birds from the biologically oriented taxonomy in the reference book, talking about "Spring birds", "Small birds", "Migratory birds", and "Birds of prey" instead of orders, families, kins etc.

- In many cases the properties of the birds were more general than the terms used in the book, for example questions about a bird's appearance, e.g. *What does a European Robin look like?* which includes plumage, size, body shape, description of beak and feet, etc.

Since the two conceptualisations had many objects and properties in common and these were related in similar ways they could be integrated in the following way (cf. figure 2). Taking the system-oriented ontology as a starting point the new categories of birds found in the question corpora were added. Allowing multiple inheritance new links between existing categories and new categories were added. Note, for example, how the new category "Small bird" is introduced and a new link is added to "Finches" in figure 2. In a similar manner the vague properties were introduced and linked to the existing properties. This is illustrated in figure 2 where two new levels are introduced, "Wingspan" and "Length" are sub-properties of the property "Size", which in turn is a sub-property of the property "Appearance".

4 Evaluating BIRDQUEST

As stated above BIRDQUEST was developed based on a corpus of questions. For further development of BIRDQUEST, we needed to assess its strengths and limitations during dialogues with real users. An evaluation of the system was thus performed with

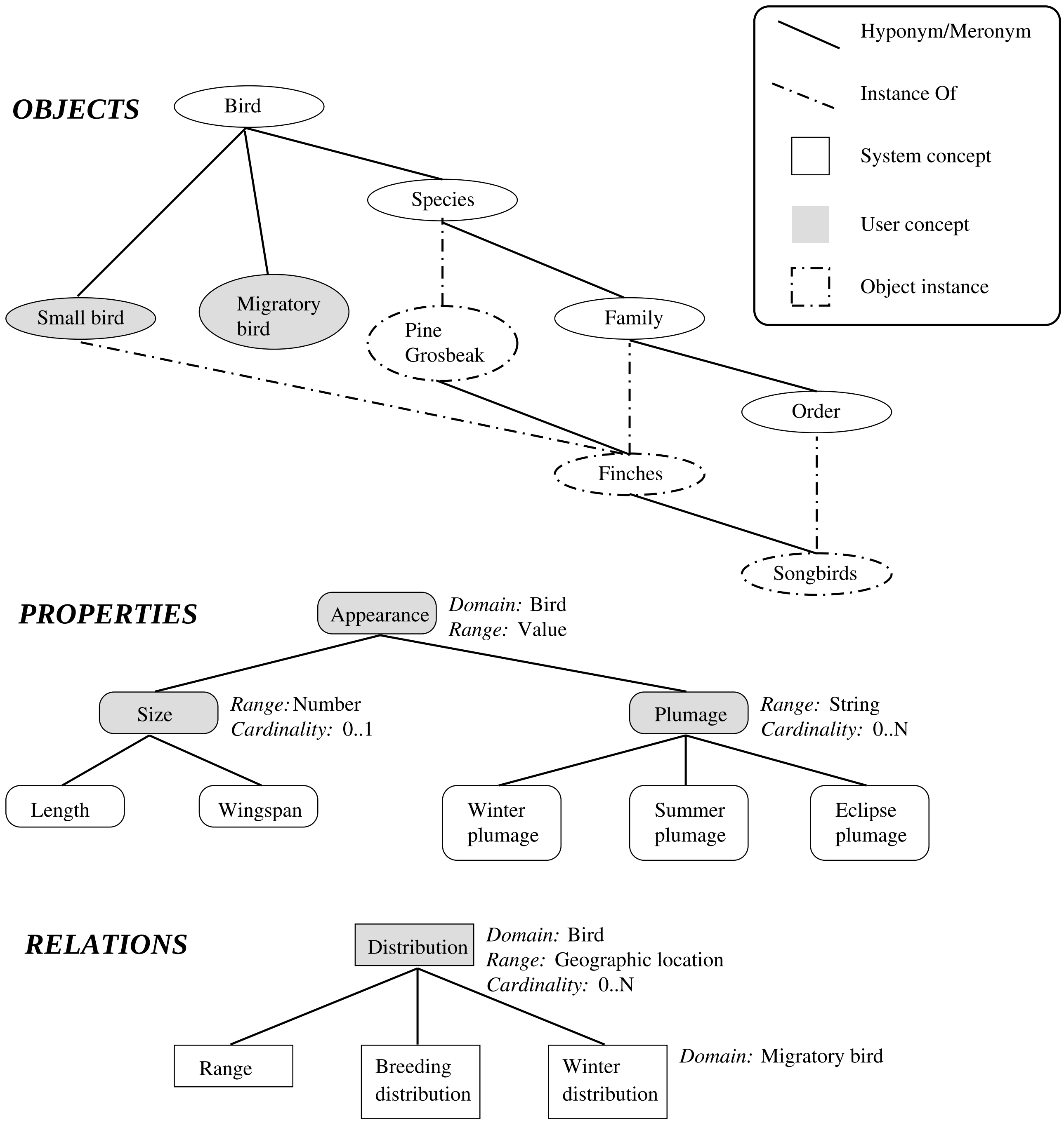

Figure 2: A part of the integrated ontology representing the conceptualisations of both bird encyclopaedia and users.

the goal of detecting problems concerning interpretation, dialogue management, and representation and use of domain knowledge.

4.1 Data collection

BIRDQUEST is intended to be used by casual users without previous experience of dialogue systems or extensive knowledge of birds. It was therefore evaluated in a walk-up and use situation similar to a real use situation during a day when the public was invited to the university. In that respect the situation resembles that of Gustafson and Bell (2000), though slightly more controlled.

We had six machines running BIRDQUEST during 2 hours and 30 minutes and collected dialogues from 27 users. They received minimal instructions in advance, they were only told that the system can answer questions on Nordic birds, that it understands Swedish, and that the dialogue would be recorded.

The resulting corpus consisting of 27 dialogues have a total number of 518 user utterances, with a mean of 19 for each user. However, with individual differences, for instance, three users posing more than 40 utterances to the system and three users posing less than 5.

Personal data about age, gender, interest in birds, and knowledge of birds were collected together with each dialogue. The users where of varying age, 5 female and 22 male. Most of them had no interest in birds, nor any knowledge of birds. Thus, despite having no interest in birds, they were fairly representative of the intended users. Besides the logged dialogue, the users were also asked to fill out a small questionnaire on how they liked to use the system. Most users thought the system was fun to use, on a 10-graded scale we had a mean of 7.1. The users also though that it was fairly easy to use BIRDQUEST, mean 6.1. On the question how they liked the system we had a score of 4.7, i.e. the users neither disliked nor liked BIRDQUEST.

4.2 Corpus annotation and initial analysis

As we had no predefined tasks we did not have a situation that allowed for a controlled evaluation, as e.g. PARADISE (Walker et al., 1998) or PROMISE (Beringer et al., 2002). Instead we used a combination of quantitative and qualitative approaches to analyse the collected dialogue corpus. The dialogues were tagged in order to provide statistics over successful and problematic information exchanges.

The user utterances were categorised as in Table 1 and the categorisation of responses from BIRDQUEST is presented in Table 2.

Table 1 shows that approximately half of the users utterances (48%) were involved in successful information exchanges where the user initiated an information request or answered a clarification request from the system. We also see that 25% of the user´s utterances are erroneous in some way and that BIRDQUEST failed on 141 utterances, as will be further discussed in section 5.

Table 1: User utterances

	No of utterances	Percentage of user utterances
Interpretable Requests	189	37%
Cooperative CR Responses	55	11%
Uncooperative CR responses	11	2%
Out of scope	121	23%
Mis-interpreted	141	27%

Table 2: System utterances

	No of utterances	Percentage of system utterances
Successful resp.	180	35%
Clarification req.	70	13%
Incorrect resp.	15	3%
Incorrect focus	16	3%
Error message	240	46%

From Table 2 we see that BIRDQUEST presented 180 successful responses. A successful response is a response where BIRDQUEST presents information found in the database. A response where the bird encyclopaedia does not include the information and BIRDQUEST responds e.g. *Information on wing span is missing for magpie.* is, however, also considered successful. The reason being that BIRDQUEST successfully accessed the database and presented whatever information was there, including cases where there was no information in the database. Among the 180 there are 55 such responses, so they are not rare, and shows one of the many interesting problems we encountered in the development of a dialogue system based on information extraction from a text book.

4.2.1 Clarifications

Notable is that a fair amount of the dialogue moves involve clarifications. The system initiates 70 clarification sub-dialogues in order to transform a vague information request to a specific question, as exemplified by the excerpt below [1].

U17:	What do water fowls look like?
S18:	Choose one of the following Water fowls: Barnacle Goose, Tufted Duck, Bean Goose, Whooper Swan, Gadwall,...
U19:	gadwall
S20:	You can receive information about the size and plumage of a Gadwall. Please specify what you are interested in.
U21:	size
S22:	A Gadwall has a length of 56 cm from beak tip to tail. Information about wingspan is missing for Gadwall

The basis for this type of clarification is domain knowledge collected from the domain ontology. Utterance U17 is under specified since the object, the bird family ”Water fowls”, can refer to a number of different species of birds, and the property ”Appearance”, is vague. To pose clarification question S18, information about which species belong to the given family is gathered from the ontology and the user is asked to chose one of them. Next, in S20, the ontology is accessed to retrieve the sub-properties of appearance. When the user has chosen a specific property (U21) the request is sufficiently specified. The ontology is used to find the sub-properties of ”Size” and these are then used to access the database and the result is presented to the user (S22).

The users responded cooperatively to 55 clarification requests from the system and incorrectly 11 times. A typical example of the latter is seen below.

S22:	You can receive information about size and plumage of a Blue Tit. Please specify what you are interested in.
U23:	blue tit

Dialogue management, such as clarification subdialogues, thus plays an important role for the performance of BIRDQUEST.

Contextual interpretation and dialogue history management are other important dialogue phenomena from MALIN that are frequently utilised in the dialogues. Managing dialogue history is, however, not trivial. There are 16 cases in the corpus, termed Incorrect focus in Table 2, when BIRDQUEST presents doubtful responses because of how dialogue history is handled, as will be further discussed in section 5.1.

[1] All examples are translations of excerpts from the Swedish dialogue corpus.

4.2.2 Utterances out of scope for BIRDQUEST

Approximately half of the non-successful user utterances (23% of all user utterances) were questions that BIRDQUEST will never be able to answer. Beringer et al. (2002) use the term incooperative user for users who ”fall out of the role or purposely misuse the system.”, and propose to exclude them in evaluations. We include such users in our corpus, but group them together in a wider category called Out of scope.

Out of Scope utterances include user requests for information that is outside the scope of the application, such as *How do you kill crows?*, or socialisation utterances (Gustafson and Bell, 2000) such as *How are you?*. Utterances can also be out of the database' scope, e.g. *How high does a magpie fly?* is such an utterance since there is no information on how high birds fly in the Bird encyclopaedia. These type of requests are further discussed in section 5.5

The reason for grouping such utterances together is that BIRDQUEST can never present information to them. Instead, we need to add a number of well-designed responses informing the user on the system's abilities. Utterances that are out of the system's scope require different types of responses from the system, and the corpus gave us valuable insights on the importance of system help messages describing what BIRDQUEST can and cannot do.

4.2.3 Utterances where BIRDQUEST fails

Finally, there are those utterances where the system failed, i.e. those where an answer can be found in the encyclopaedia, but where BIRDQUEST fails to present a successful response for various reasons. Such utterances comprise 27% of the users' input.

We have further analysed these and categorised them as being 1) spelling mistakes, 2) lexical gaps, or 3) grammatically out of scope, as seen in Table 3. Table 3 includes only utterances that can be successfully responded to, not, for instance, misspellings in utterances that are out of the systems' scope.

Table 3 only gives a very brief indication on the nature of non-interpretable utterances in the corpus. For instance, each utterance is tagged as being of one type only, with misspellings having highest priority and missing grammar rules the lowest. Furthermore, there could be several misspellings in one utterance.

It is also the case that the categories overlap, i.e.

Table 3: User utterances not interpreted by BIRDQUEST

	No of utterances	Percentage of system utterances
Misspelled	28	5%
Not in lexicon	64	12%
Not in grammar	50	10%

utterances can belong to more than one category, e.g. a misspelled word can also be missed in the lexicon. There are three such utterances tagged misspelled that also contain words not in the lexicon. 11 of the utterances tagged misspelled cannot be handled because of missing grammar rules. Thus, with a spelling checker, BIRDQUEST can handle another 14 utterances, i.e. misspelling on its own only accounts for 2,5% of the utterances that BIRDQUEST cannot handle.

5 Implications for development

Our main goal was to collect data for further development. Utterances in the categories involving unsuccessful information exchanges, such as incorrect responses and mis-interpretations, were therefore analysed more closely.

5.1 Problematic focus management

Focus management and methods to determine topic shifts play an important role in many dialogue systems. Our corpus includes a variety of problematic cases, that need further consideration, both on how focus is handled in the dialogue manager and how the grammar rules are constructed.

When a user poses a question that gives rise to a partially interpreted request the dialogue manager seeks the dialogue history to see if it can be contextually specified. This works well in most cases when utterances are elliptic but in some cases, especially when the interpreter has failed to produce a complete interpretation, it can cause faulty responses, see the example below.

U35: How old does a common gull get?
S36: Information about the greatest age of a Common Gull:
Greatest age: 29 years and 7 months
U37: What migratory birds exist?
S38: Information about range is missing for Common Gull.

The reason for not being able to correctly interpreting U37 is that migratory birds is not a well defined concept in the bird encyclopaedia [2]. Thus, only a partial interpretation with the property ”Range”, based on interpreting ”exist” as connected to a location, is produced. The partial interpretation is then combined with the object ”Common gull” from U35 and an incorrect response is produced in S38.

A more difficult problem to handle is when there has been a long segment of uninterpreted user utterances often after sequences of misspellings or requests outside the scope of the system. Normally, such sequences of utterances are contextually interpreted utilising the dialogue history and a successful response can be presented to the user. There are, however, also ”faulty” cases, as seen in the dialogue fragment below[3]. Here, the user is trying to shift focus and thus the property ”Feed” should not have been inherited from the previous question (U11). A more appropriate response in S18 would have been *You can receive information about a song thrush's flight, feed, breeding location, appearance,... Please specify what you are interested in.*

U11: What do the honey buzzard eat
S12: Information about feed for a Honey Buzzard:
Main feed consists of Honeycomb,[...]
U13: What birds can imitate speech
S14: Failed to interpret the question. Please try again
U15: Talking birds
S16: Failed to interpret the question. Please try again
U17: Song thrush
S18: Information about feed for a Song Thrush:
Worms, larvas and mollusc etc[...]

The problem of dialogue history and focus shifts does not have a simple solution. Leceuche et al. (2000) use focus rules based on relations between the various things in the domain to track focus shifts. A similar approach is to encode the domain in a type hierarchy and use underspecified feature structures to determine clarification requests (Denecke, 1997). Jokinen et al. (1998) use a topic model based on a tree organisation of domain knowledge to detect topic shifts. Such approaches would correctly handle the dialogue above and start a new focus space after U13, as there is no relation

[2]Many bird species do not migrate every year and often some exemplars stay when most birds of that specie migrate etc. cf. section 5.4.

[3]The Swedish name of Song thrush, (Sw. taltrast), can be literally translated as ”Talking thrush”

between the items in U11-S12 and U13. However, as we assume that it is a good idea to keep items in focus even after user utterances with no previous relation to the current focus (Jönsson, 1995), such mechanisms are not applicable here. For instance, if the user instead typed ”U13b: magpi”, we would not have a relation between the items in the utterances, and can then not be able to correctly respond if the user corrected that to ”U15b: magpie”.

Another approach is proposed by Chu-Carroll (2000) who distinguish between task initiatives and discourse initiatives that are identified based on cues. These are then utilised to adopt different interaction strategies. In BIRDQUEST one solution could be to keep focus for discourse initiatives and drop focus for new task initiatives. In the dialogue above, U13, is to be considered a new task initiative whereas U13b is not a new task initiative and focus is maintained.

5.2 Unnecessary clarifications

Since information extraction is used to populate the database with information there can be holes in it, i.e. for some properties values are missing due to shortcomings in the information extraction component or lack of information in the original text source. This causes unnecessary clarifications. In the example below a more appropriate answer to U19 would have been to give information about size instead of making a clarification since there is no information about plumage in the database.

U19: What does a mute swan look like?
S20: You can receive information about the size and plumage of a Mute Swan. Please specify what you are interested in.
U21: What do the plumage of a mute swan look like?
S22: Information about eclipse plumage is missing for Mute Swan.
Information about winter plumage is missing for Mute Swan.
Information about summer plumage is missing for Mute Swan.
Information about breeding plumage is missing for Mute Swan.

The database has many such missing values as very often there was no information in the information source. It is, thus, not possible to adopt general dialogue strategies, for instance, for clarification requests. Instead, the domain knowledge base must be consulted to decide on a proper dialogue strategy for each property.

In BIRDQUEST the unnecessary clarifications can be dealt with through extended co-operation between the dialogue manager and the domain knowledge manager. When a vague property is encountered the dialogue manager can send the request to the domain knowledge manager for information on suitable clarifications. By traversal of the ontology the DKM can produce a set of sub-properties which can be used to access the database. Should there be only a few pieces of information these can be returned directly as an answer to the request. Otherwise the proper clarification is decided based on the relation between the given property and the sub-properties which produced the values.

Note, however, that such a strategy can violate the learning aspect discussed below.

5.3 Partial and empty answers

A problem related to unnecessary clarifications are how partial and empty answers should be presented to the user when a vague property has been mapped to several sub-properties, for example ”Plumage” in S22 in the example above, or ”Size” in the example below.

S2: You can receive information about the size and plumage of a Magpie. Please specify what you are interested in.
U3: size
S4: Information about wingspan is missing for Magpie.
A Magpie has a length of 46 cm from beak tip to tail.

In the case of empty answers, S22 above, an alternative could be to just state that *Information about plumage is missing for mute swan*, and for partial answers the sub-properties with missing information could be omitted. However, including all the sub-properties have a point in so far as it helps the user learn what type of information the system can provide. A solution is to make a compromise and make the system adaptive. This means that the first time a vague property is mapped to a set of sub-properties all of them are presented regardless of whether they have values or not, thereafter empty answers can be omitted.

5.4 Ontological interpretation failures

Some seemingly simple interpretation failures that have been tagged as missing information in the lexicon are actually more complicated and relates to the representation of domain knowledge in the ontology. For example, users tend to ask what colour a specific bird have, see U1 below[4]. However, in the bird encyclopaedia colour is linked to the body parts of a bird. The complex relation between the concept "Bird" and the property "Colour" via the body parts thus have to be captured in the ontology so that this knowledge can be used for interpretation and dialogue management.

U1: What colour is a blue tit?
S2: Failed to interpret the question. Please try again

With more information about how concepts and properties are related, for example, through a chain of hypernym and meronym relations, requests like U1 could be dealt with. Since a bird consists of body parts and body parts have colours the missing link in the request is a body part and based on this a clarification could be produced *A bird has several colours for different body parts. What part of the blue tit are you interested in?*

Migratory birds, as discussed in section 5.1, is another example when the ontology can be used to stipulate a certain interpretation. With the help of an ornithologist, we can define what migrating behaviour constitutes a migratory bird and add that to the ontology. Used by the information extraction component, we can easily pick those birds that are migratory birds, and as we utilise the same ontology for the interaction component, we can interpret such requests. Note, however, that this goes beyond our initial goal of only utilising the bird encyclopaedia.

5.5 Questions outside database coverage

Users often posed questions that were outside the database coverage. A successful interpretation of such a request and an appropriate response is seen in the dialogue below:

U1: What is the wingspan of an Albatross?
S2: There is no information on non-Nordic birds. The available birds are listed in the menu "Birds a-ö". Please try again

The same strategy could be used to handle similar questions concerning non-Nordic birds like canaries and ostriches. However, that would require addition of information of non-Nordic birds to the ontology collected from other sources than the bird encyclopedia.

However, in most cases the requests concerned properties that are not covered by the database but users often ask about, for example "Weight" and "Flight-speed" as in *How much does a Sea Gull weigh?* or *How fast can an Eagle fly?*. The common response to these types of questions were *Failed to interpret the question. Please try again* or in some cases a partial interpretation was made which led to inappropriate responses. A more desirable response would be to give more informative error messages and explain to the user that it cannot answer questions about these topics.

Extending the ontology could help give informative answers when the questions are outside database coverage. The properties similar to those in the database, such as "Weight", "Flight-speed", could be added to the ontology as user-oriented properties. Since the DKM always have to map this type of properties to the system-oriented sub-properties before database access it could conclude that, if a user-oriented property do not have any user-oriented sub-properties, it is outside database coverage and an appropriate answer can be given. If these properties were related to others, for example, "Weight" is a sub-property of "Appearance", the system could even suggest some of the sibling properties, in this case "Size" and "Plumage".

Another strategy is to have BIRDQUEST respond with help phrases explaining how to pose valid requests, as is done in Targeted Help (Gorrell et al., 2002). Targeted help is used for improving user behaviour in speech interfaces. It utilises the SLM-based recognition and categorised help message templates to present targeted help when the grammar based recogniser fails. Thus, a system must learn the most common types of mistakes which in turn must be classified to provide a targeted help. Unfortunately, we do not yet have a large enough BIRDQUEST corpus for such classification.

6 Summary

In this paper we have presented an evaluation of a dialogue system that was developed to access a database built from information automatically extracted from a text book. The results from our eval-

[4]Many of the requests for appearance can be handled by presenting a picture of the bird. However, the pictures in our bird encyclopaedia are copyrighted and can therefore not be presented.

uation show that it is possible to develop such a system and that users staying within the boundaries of the application will get useful information.

Dialogue is important for the interaction as well as a shared ontology for both information extraction and interaction. The evaluation also revealed a number of challenging issues, especially regarding, system help messages, dialogue management, problems with gaps in the database due to incomplete information and how to utilise a domain ontology.

7 Acknowledgement

Many thanks to Frida Andén, Lars Degerstedt and Sara Norberg for interesting discussions and work on the development and implementation of the BIRDQUEST system. This research is financed by Vinnova, Swedish Agency for Innovation Systems

References

Jan Alexandersson and Norbert Reithinger. 1995. Designing the dialogue component in a speech translation system. In *Proceedings of the Ninth Twente Workshop on Language Technology (TWLT-9)*, pages 35–43.

Nicole Beringer, Ute Kartal, Katerina Louka, Florian Schiel, and Uli Türk. 2002. Promise - a procedure for multimodal interactive system evaluation. In *Proceedings of the Workshop 'Multimodal Resources and Multimodal Systems Evaluation'. Las Palmas, Gran Canaria, Spain*.

J. Burger, C. Cardie, V. Chaudhri, R. Gaizauskas, S. Harabagiu, D. Israel, C. Jacquemin, C. Y. Lin, S. Maiorano, G. Miller, D. Moldovan, B. Ogden, J. Prager, E. Riloff, A. Singhal, R. Shrihari, T. Strzalkowski, E. Voorhees, and R. Weishedel. 2001. Issues, tasks and program structures to roadmap research in question & answering (Q&A). http://wwwnlpir.nist.gov/projects/duc/papers/qa. Roadmap-paper_v2.doc.

Jennifer Chu-Carroll. 2000. MIMIC: An adaptive mixed initiative spoken dialogue system for information queries. In *Proceedings of 6th Applied Natural Language Processing Conference*, pages 97–104.

Nils Dahlbäck, Annika Flycht-Eriksson, Arne Jönsson, and Pernilla Qvarfordt. 1999. An architecture for multi-modal natural dialogue systems. In *Proceedings of ESCA Tutorial and Research Workshop (ETRW) on Interactive Dialogue in Multi-Modal Systems, Germany*.

Matthias Denecke. 1997. An information-based approach for guiding multi-modal human-computer-interaction. In *IJCAI´97, Nagoya, Japan*, pages 1036–1041.

Annika Flycht-Eriksson and Arne Jönsson. 2000. Dialogue and domain knowledge management in dialogue systems. In *1st SIGdial Workshop on Discourse and Dialogue, Hong Kong*.

Genevieve Gorrell, Ian Lewin, and Manny Rainer. 2002. Adding intelligent help to mixed initiative spoken dialogue systems. In *Proceedings of ICSLP 2002*.

Tom R. Gruber. 1993. A translation approach to portable ontology specification. *Knowledge Acquisition*, 5:199–220.

Joakim Gustafson and Linda Bell. 2000. Speech technology on trial: Experiences from the august system. *Natural Language Engineering*, 6(3-4):273–286.

Eli Hagen. 1999. An approach to mixed initiative spoken information retrieval dialogue. *User modeling and User-Adapted Interaction*, 9(1-2):167–213.

Arne Jönsson and Magnus Merkel. 2003. Some issues in dialogue-based question-answering. In *Working Notes from AAAI Spring Symposium, Stanford*.

Kristiina Jokinen, Hideki Tanaka, and Akio Yokoo. 1998. Context management with topics for spoken dialogue systems. In *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics and 17th International Conference on Computational Linguistics, COLING-ACL'98, Montreal*, pages 631–637.

Arne Jönsson. 1995. Dialogue actions for natural language interfaces. In *Proceedings of IJCAI-95, Montréal, Canada*.

Renaud Leceuche, Dave Robertson, Catherine Barry, and Chris Mellish. 2000. Evaluating focus theories for dialogue management. *International Journal on Human-Computer Studies*, 52:23–76.

Kavi Mahesh and Sergei Nirenburg. 1995. A situated ontology for practical NLP. In *Proceedings of IJCA'95 Workshop on Basic Ontological Issues in Knowledge Sharing*, Montreal, Canada.

Dan Sullivan. 2001. *Document Warehousing and Text Mining*. John Wiley & Sons.

Marilyn A. Walker, Diane J. Litman, Candace A. Kamm, and Alicia Abella. 1998. Paradise: A framework for evaluating spoken dialogue agents. In Mark Maybury & Wolfgang Wahlster, editor, *Readings in Intelligent User Interfaces*. Morgan Kaufmann.