

WikiTalk: A Spoken Wikipedia-based Open-Domain Knowledge Access System

Graham WILCOCK

UNIVERSITY OF HELSINKI, Finland
graham.wilcock@helsinki.fi

ABSTRACT

WikiTalk is an open-domain knowledge access system that talks about topics using Wikipedia articles as its knowledge source. Based on Constructive Dialogue Modelling theory, WikiTalk exploits the concepts of Topic and NewInfo to manage topic-tracking and topic-shifting. As the currently talked-about Topic can be any Wikipedia topic, the system is truly open-domain. NewInfos, the pieces of new information to be conveyed to the partner, are associated with hyperlinks extracted from the Wikipedia texts. Using these hyperlinks the system can change topics smoothly according to the user's changing interests. As well as user-initiated topics, the system can suggest new topics using for example the daily "Did you know?" items in Wikipedia. WikiTalk can be employed in different environments. It has been demonstrated on Windows, with an open-source robotics simulator, and with the Aldebaran Nao humanoid robot.

KEYWORDS: Open-domain knowledge access, spoken dialogue system, Wikipedia, human-robot interaction.

1 Introduction

The paper describes WikiTalk, an open-domain knowledge access system that talks about topics using Wikipedia articles as its knowledge source. The system is truly open-domain in the sense that the currently talked-about topic can be any topic that Wikipedia has an article about, and the user can switch topics at any time and as often as desired.

In an open-domain system it is extremely important to keep track of the current topic and to have smooth mechanisms for changing to new topics. In WikiTalk, topic-tracking and topic-shifting are managed with the help of the concepts of Topic and NewInfo, in accordance with the Constructive Dialogue Modelling theory of Jokinen (2009). The distinctive contribution of WikiTalk is that NewInfos, the pieces of new information to be conveyed to the partner, are associated with hyperlinks extracted from the Wikipedia texts. Using these hyperlinks the system can change topics smoothly according to the human's changing interests.

Interaction technology has to address the engagement of the user in the interaction. The system has to manage the interaction so that there is a natural conversation rather than a monologue on a particular topic. For instance, in teaching and learning situations such conversational capability is important. This requires dynamic tracking not only of dialogue topics but also of the users' focus of attention and of the user's level of interest in the topic. Techniques for attention-tracking and interest-tracking in interactive situations are important parts of the system.

WikiTalk has been developed so that it can be integrated into different hardware and software environments. In spoken human-computer interaction scenarios, such as a hands-free in-car conversational companion, WikiTalk requires suitable speech recognition and speech synthesis components. The system can be used on Windows, using the standard Windows Speech Engine components. In embodied agent scenarios, for example in human-robot interaction, the system needs to be integrated with appropriate multimodal components for face-tracking, nodding and gesturing, proximity recognition and so on.

An extended version of WikiTalk has been implemented on the Aldebaran Nao robot, as the basis for multimodal human-robot conversational interaction (Jokinen and Wilcock, 2012b; Csapo et al., 2012). We are not aware of any previously-reported multimodal human-robot conversational interaction system that is open-domain.

The paper is structured as follows. Section 2 presents background on dialogue modelling and human-robot interaction. Section 3 introduces the basic approach to implementing open-domain dialogues using Wikipedia. Section 4 describes how smooth topic shifts are handled. Section 5 presents an example open-domain conversation with a robot. Section 6 considers related work and Section 7 describes multimodal extensions for humanoid robots.

2 Dialogue Modelling and Human-Robot Interaction

The theoretical basis of WikiTalk is Constructive Dialogue Modelling (Jokinen, 2009), which integrates topic management, information flow, and the construction of shared knowledge in the conversation by communicative agents. We have applied this model to human-robot interaction, in which cooperation manifests itself in the system properties that allow users to interact in a natural manner, i.e. in the ways in which the system affords cooperative interaction. The agents' goals can range from rather vague "keep the channel open"-type social goals to more specific, task-oriented goals such as planning a trip, providing information, or giving instructions.

Jokinen and Wilcock (2011) describe different levels of emergent verbal behaviour that arise when speech is used in human-robot interaction. At the first levels of verbal behaviour the robot produces spoken monologues giving a stream of simple explanations of its movements, and the human uses voice commands to direct the robot's movements. At the next level, cooperative verbal behaviour begins to emerge when the robot modifies its own verbal behaviour in response to being asked by the human to talk less or more.

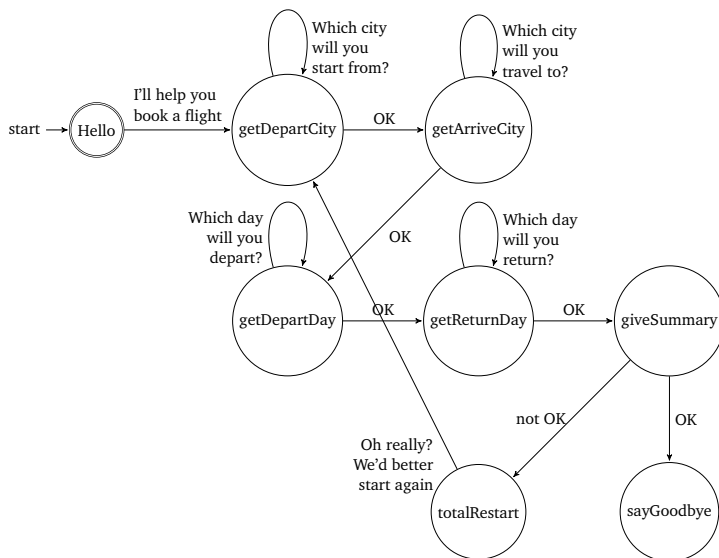


Figure 1: An example finite-state machine for a closed-domain dialogue.

2.1 Closed-Domain Dialogues

Moving up a level, the robot asks questions in order to achieve a specified dialogue goal. The classic example of this kind of dialogue is a flight reservation system. Finite state machines have been successfully used for these closed-domain form-filling dialogues. An example is shown in Figure 1.

Here, the system asks questions in order to achieve the dialogue goal by filling in the required fields in the form. The specific questions are predefined for each specific domain. Although it is easy to change the details about the flights, the destinations and the departure times that are maintained in the system's database, it is very difficult to change to a different domain.

In order to move up another level, to open-domain dialogues, we use Wikipedia as a source of world knowledge. By exploiting ready-made paragraphs and sentences from Wikipedia, WikiTalk enables a robot to talk about an almost infinite range of topics. The robot can also perform smooth topic-shifts according to the human's interests, by using hyperlinks extracted

from the Wikipedia articles. As well as human-initiated topics, the robot can suggest new topics using for example the daily "Did you know?" items in Wikipedia.

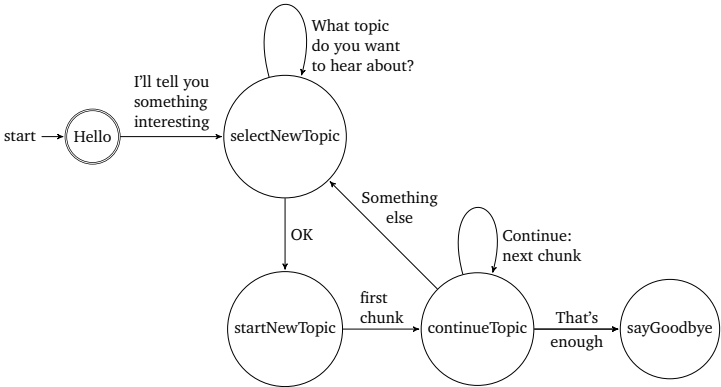


Figure 2: Towards a finite-state machine for open-domain dialogues.

3 Getting Started with Open-Domain Dialogues

In order to support open-domain dialogues, the dialogue control model in WikiTalk is different from the closed-domain form-filling systems. The basic interaction management is still controlled by finite-state transitions, but the set of states are not linked to specific domain-related items such as "get departure day" or "get destination city" in a closed-domain system. Rather, the states are used to manage the flow of the interaction with the user, especially topic-tracking and topic-shifting. This approach makes it possible for a finite number of states to manage dialogues with an infinite number of topics.

The following sections present an introduction to some of the states needed for open-domain dialogues, showing how conversations can be started and stopped, and how topics within a conversation can be started, continued and stopped. A partial diagram of these states is shown in Figure 2

3.1 Hello and Goodbye

Like closed-domain dialogue systems, an open-domain system needs a way to get started and a way to finish. These are handled by a Hello state and a Goodbye state. As well as saying "Hello" and "Goodbye", these states can be extended with suitable behaviours. For example, a mobile robot can stop moving around for the duration of the conversation. A humanoid robot can stand up to talk, or sit down to talk, depending on the scenario.

3.2 Selecting a Topic

After the Hello state, the system moves to the SelectTopic state. There are different ways to select a topic. The easiest way to get started is to select a topic from a list of favourites or from

a history list of recently talked-about topics. WikiTalk provides convenient ways to extend the favourites by adding the current topic to the list, and to remind the user about the favourites or the recent history topics. Other ways to select a topic are described in Sections 4.6 and 4.7.

3.3 Starting a New Topic

When a new topic has been selected, the `NewTopic` state gets the text for the topic from Wikipedia. As Wikipedia articles are designed for visual inspection in web browsers, with visual layouts, footnotes, and various special symbols, a certain amount of cleaning-up and reformatting is necessary to make the text suitable for reading by a speech synthesizer.

In addition, the text is divided into chunks (paragraphs and sentences) of a suitable size for spoken dialogue contributions. The appropriate chunk size depends on several factors. If the system has a good interrupt mechanism (Section 3.5), a large chunk size (whole paragraphs) is fine as the user can easily stop the system talking at any point. Otherwise, a small chunk size (a single sentence) is better, so the system can check frequently between chunks for positive or negative feedback before continuing or stopping.

3.4 Continuing the Same Topic

Once a new topic has been started, the `ContinueTopic` state manages the system's presentation of the chunks of the topic text. If the user asks for more, or otherwise shows some interest in the topic, the system continues with the next chunk. If the user keeps asking for more until the end of the article is reached, the system says that there is no more information about the current topic and moves to the `SelectTopic` state (Section 3.2) which asks the user to select a new topic.

At the end of each chunk about a given topic, the user can ask for the same chunk to be repeated, or can ask to go back to the previous chunk about the current topic. The user can also interrupt the current chunk without listening to it all, and ask to skip forward to the next chunk on the same topic.

3.5 Interrupting the System

It is helpful if there is a simple way to interrupt the system while it is talking. Otherwise, the user has to wait until the system finishes the current chunk before telling it to stop. This can be annoying if the system has mistakenly chosen the wrong topic and the chunk size is a whole paragraph.

The best interrupt mechanism depends on the hardware and software environment. Some speech engines support barge-in, others do not. Some robots have a convenient button or sensor that can be used to interrupt the robot's behaviour. In this case, large paragraph-size chunks can be interrupted whenever desired.

When the system is interrupted, it stops talking and moves to an `Interrupt` state, remembering which state it was in when the interruption occurred, what the topic was, and which chunk it had reached. It explicitly acknowledges the interruption by saying "Oh sorry!" and waiting for the user's input. The user can then tell it to continue, to go back to an earlier chunk, to skip forward to the next chunk, or to switch to a new topic.

4 Enhancements for Open-Domain Dialogues

In natural human-human dialogues, the topic changes dynamically as the conversation goes along. When the topic changes smoothly from the current topic to a related topic, there is no need to make the change of topic explicit. A change to a related topic is normally a smooth topic shift, and this occurs more or less continuously. It is important that dialogue systems have smooth mechanisms to support smooth topic shifts.

4.1 Topic Trees

Previously, dialogue systems have often used topic trees to organize knowledge into related topics. Focus trees (McCoy and Cheng, 1991) were originally proposed to trace foci in natural language generation systems. The branches of the tree show what sort of shifts are cognitively easy to process and can be expected to occur in dialogues. Random jumps from one branch to another are unlikely, and if they do occur, they should be appropriately marked. The tree both constrains and enables prediction of what is likely to be talked about next.

Our approach finds an equivalent to topic trees by exploiting the hyperlinks found in Wikipedia. These links provide a ready-made organisation of domain knowledge, for almost any domain. We believe this approach is better than hand-coding topic trees or automatic clustering. Kirschner (2007) gives an overview of these approaches to Interactive Question Answering. Jokinen et al. (1998) combine a manually built tree for main topics with an n-gram model for topic shifts. Kirschner and Bernardi (2009) use machine learning to explore follow-up questions.

Instead of attempting a deep processing approach involving information extraction, question answering or summarization techniques, we prefer a shallow processing approach in which selected chunks of the Wikipedia texts are read out aloud, with a relatively small amount of reformatting and clean-up necessary for spoken contributions. This shallow approach allows us to concentrate on identifying hyperlinks and on managing the topic shifts smoothly.

4.2 Smooth Topic Shifts

When the system is talking about the current topic, the chunks contain NewInfos, pieces of new information about the current topic. In Wikipedia, these NewInfos are typically annotated with hyperlinks to articles about the related topic. WikiTalk analyses the texts of the Wikipedia articles that it reads, and extracts the hyperlinks that are included in the texts.

In order to switch to a new topic that is related to the current topic, the user just says the name of the NewInfo that is interesting. For example, if the system is talking about Shakespeare and says "Shakespeare was born in Stratford-upon-Avon", the user can say "Stratford-upon-Avon?" and the system will smoothly switch topics and start talking about Stratford-upon-Avon. It does this by going to the NewTopic state (Section 3.3) with Stratford-upon-Avon as the new Topic. This important state transition, from the ContinueTopic state to the NewTopic state, is shown in Figure 3. It is WikiTalk's smooth mechanism for handling smooth topic shifts.

These smooth topic shifts can only be performed when the relevant NewInfo is marked-up with a hyperlink in the Wikipedia text. This is not normally a problem, because the authors of the Wikipedia texts typically provide suitable links, and if the authors don't provide them the links are usually added later by readers and editors. This is one of the reasons why Wikipedia is so convenient and so widely-used.

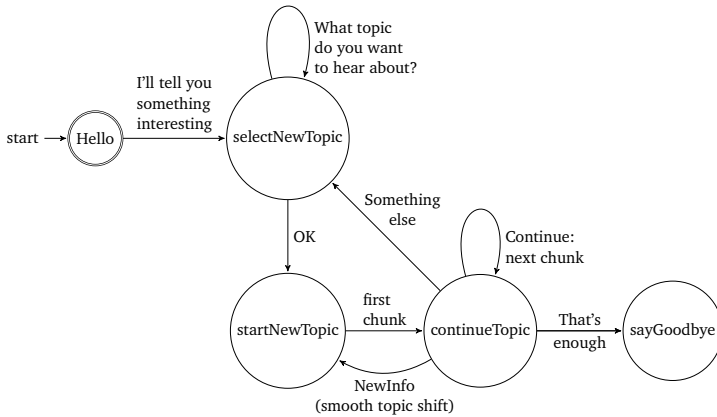


Figure 3: Adding smooth topic shifts.

These hyperlinks were inserted into the texts by the human authors precisely because they represent cognitive links between the concepts in the current text and other concepts in other texts. These hyperlinked NewInfos are precisely the related topics that the user is most likely to want to know more about next. In a normal web browser the user clicks on one of these links in order to pursue a particular concept or topic that may be of greater interest than the topic of the current article. In Wikitalk, the user says one of the link words for the same reason, to move to a new topic in order to pursue a particular concept that may be of greater interest.

4.3 Hyperlinks in Wikipedia

WikiTalk is able to take advantage of the presence of these carefully-chosen links in the Wikipedia articles. However, there are two possible problems that can arise.

The first problem occurs when hyperlinks are added to the first occurrence of a word in an article, but are not added to every subsequent occurrence of the same word in the rest of the article. This problem can be solved by extending the list of words to be recognized, to include all the previously hyperlinked words in the article, not only the hyperlinked words in the latest spoken chunk.

The second problem is that the number of topics in Wikipedia is constantly growing, and therefore more and more topic links are available for linking to more and more words in the articles. The fear is that eventually every word will be hyperlinked. So far this has not happened. However, if every word were hyperlinked, the speech recognizer in WikiTalk would in effect be trying to do open vocabulary speech recognition.

4.4 Changing-Vocabulary Speech Recognition

For effective speech recognition, it is very important to limit the vocabulary to be recognised as much as possible. Open vocabulary speech recognition is currently not feasible as there are simply too many possible words (maybe 300,000 different words in English). An attempt to develop an open-vocabulary speech interface to the Google search engine is described by Franz and Milch (2002).

In WikiTalk, the speech recognizer listens for a relatively small number of the extracted link words, as well as a small number of commands. The commands allow the user to control the system behaviour in order to start, continue, go back, repeat, stop and so on. The list of commands is relatively small and relatively fixed. The list of extracted link words is constantly changing, and its length varies depending on the number of links in each article, but it is nevertheless relatively small. This type of changing-vocabulary speech recognition is naturally far more effective than open-vocabulary recognition.

4.5 Speech Recognition Confidence Scores

WikiTalk uses any speech recognizer that is available, for example Windows Speech Engine or the Aldebaran Nao robot speech components. The details therefore depend on the specific environment, but word spotting is used where possible so that link words can be recognized when they form part of longer phrases such as "Tell me about X".

Speech recognition confidence scores are used where possible, to decide how to proceed. If a word or phrase is recognized with very high confidence, the system goes ahead immediately without checking. If one word or phrase X is recognized with a significantly higher confidence score than alternatives, but not a very high score, the system asks "Did you mean X?". If two words X and Y have relatively high scores, the system first asks "Did you mean X?", and if that is wrong, "Did you mean Y?" If no words or phrases have a high score, the system immediately asks the user to repeat the input.

4.6 Awkward Topic Shifts

The easiest way to select a topic of interest is to navigate to it via the NewInfo links in other topics. However, the user may wish to talk about an entirely new and unrelated topic. This is an awkward topic shift, which needs to be explicitly signalled in human-human dialogues. In WikiTalk, a completely new and unrelated topic can be specified by spelling the first few letters. To help the speech recognizer, this is done with the standard phonetic names for the letters (Alpha, Bravo, Charlie, etc.).

Wikipedia helps this method for spelling a new topic by providing searches for topics starting with any initial sequence of letters, so it is not necessary to spell the whole topic name. After the user gives the first few letters, the system gets a list from Wikipedia of topics that start with those letters. For example, after spelling "Sierra, Hotel, Alpha, Kilo", the system asks the user to select from a list including Shakespeare, Shakira, Shakespeare in Love, Shakers, etc. The system then moves to the NewTopic state (Section 3.3) with the selected topic as the new Topic.

4.7 Suggesting Topics by the System

One of the advantages of using Wikipedia is that the system can easily suggest a topic if the user has no particular topic in mind. One way to do this is based on the "Did you know?" section on

the Main Page of the English Wikipedia. This is a short list of new or recently revised articles that is changed every day.

The system selects one of the topics from this list at random and asks the user "Did you know ...?". If the user is interested, this becomes the new topic. If not, another topic is selected from the list and the user is asked again "Did you know ...?" with another topic.

There are many other possibilities for suggesting topics using Wikipedia. For example, there is the "On this day..." section on the Main Page of the English Wikipedia. Topics can be selected at random, or the user's previous interests can be taken into account.

5 An Example Conversation with a Robot

An example dialogue about Shakespeare, using WikiTalk with a robot simulator, is given by Jokinen and Wilcock (2012a). We now give a concrete example of a conversational dialogue with an Aldebaran Nao robot, from a video recording of a user testing session at the 8th International Summer Workshop on Multimodal Interfaces, Metz, 2012.

The robot starts by standing up and telling the user that its favourite topics include Shakespeare, Metz, and Paul Verlaine. The human asks the robot to talk about Metz.

ROBOT: Metz is a city in the northeast of France located at the confluence of the Moselle and the Seille rivers. Metz is the capital and the prefecture of both the Lorraine region and the Moselle department. Located near the tripoint along the junction of France, Germany, and Luxembourg, Metz forms a central place of the European Greater Region and the SaarLorLux euroregion.

Figure 4: Starting the first topic: Metz.

In Figure 4, the robot has downloaded the Wikipedia article about Metz and reads the first paragraph. NewInfo links are extracted for *France, Germany, Luxembourg, Moselle (river), Seille, prefecture, Lorraine (region), Moselle, department, tripoint, Greater Region, SaarLorLux, Euroregion*. They are added to the list of specific items that the speech recognizer will listen for and be specifically ready to recognize. However, the user does not pick out any of these NewInfos, but asks the robot to continue with the Metz topic.

USER: Continue.

ROBOT: A Celtic oppidum, an important Gallo-Roman city, the Merovingian capital of the Austrasia kingdom, the birthplace of the Carolingian dynasty, a cradle of the Gregorian chant, and one of the oldest republics of the common era in Europe, Metz has a rich 3,000-year-history. The city has been steeped in Romance culture, but has been strongly influenced by Germanic culture due to its location and history.

Figure 5: Continuing the first topic: Metz.

In Figure 5, the robot continues talking about Metz by reading the second paragraph. In this paragraph the NewInfos are *Celts, oppidum, Gallo-Roman, Merovingian, Austrasia, Carolingian dynasty, Gregorian chant, Common era, Romance languages, Culture of Germany*. They are added to the list of specific items that the speech recognizer will listen for and be specifically ready to recognize.

The user asks about "Celtic", as shown in Figure 6. This is recognized as one of the NewInfos, because the system has extracted the hyperlink to the topic *Celts* that is associated with the word "Celtic".

USER: *Celtic?*

ROBOT: *The Celts or Kelts were an ethno-linguistic group of tribal societies in Iron Age and Medieval Europe who spoke Celtic languages and had a similar culture.*

Figure 6: Starting the second topic: Celts.

As the user is interested in "Celtic", the robot does not continue reading the main article about Metz, but instead downloads a new Wikipedia article about Celts. It is a natural continuation in the dialogue that the user will select the next topic from the NewInfos, since the human is engaged in the dialogue.

The robot starts reading the first chunk about Celts as shown in Figure 6. The NewInfos here are *Ethnolinguistics*, *Tribe*, *Iron Age*, *Middle Ages*, *Europe*, *Celtic languages*. The user asks the robot to talk about Celtic languages. The robot therefore downloads a new Wikipedia article about Celtic languages, and starts reading the first paragraph as shown in Figure 7.

USER: *Celtic languages?*

ROBOT: *The Celtic or Keltic languages are descended from Proto-Celtic, or "Common Celtic"; a branch of the greater Indo-European language family. The term "Celtic" was first used to describe this language group by Edward Lhuys in 1707.*

Figure 7: Starting the third topic: Celtic languages.

The NewInfos here are *Proto-Celtic*, *Indo-European languages*, *Language Family*, *Edward Lhuys*. The user asks about Indo-European languages, and the robot downloads a new Wikipedia article about Indo-European languages, and starts reading the first paragraph as shown in Figure 8.

USER: *Indo-European?*

ROBOT: *The Indo-European languages are a family of several hundred related languages and dialects. It has about 449 languages and dialects, according to the 2005 Ethnologue estimate, about half belonging to the Indo-Aryan subbranch. It includes most major current languages of Europe, the Iranian plateau, and South Asia, and was also predominant in ancient Anatolia. With written attestations appearing since the Bronze Age, in the form of the Anatolian languages and Mycenaean Greek, the Indo-European family is significant to the field of historical linguistics as possessing the longest recorded history after the Afroasiatic family.*

USER: *Thank you.*

ROBOT: *Thank you. Goodbye.*

Figure 8: Starting the fourth topic: Indo-European languages.

Now the user has heard enough and wants to stop. Saying "Stop" would stop the robot talking about Indo-European languages, but the robot would ask what other topic it should talk about.

Saying "Thank you" tells the robot that the user does not want to hear anything else. The robot replies "Thank you. Goodbye." and sits down.

Note that the words or phrases in the text are not necessarily identical to the names of the linked topics. WikiTalk passes to the speech recognizer the words or phrases that are actually used in the text, because those are the words that the user hears. The system maintains a mapping between these spoken words or phrases and the linked topics. For example, in Figure 6 the spoken word is "Celtic" and the linked topic is *Celts*).

6 Related work

The most famous open-domain dialogue system is still ELIZA (Weizenbaum, 1966). Of course, the reason that ELIZA was capable of maintaining an on-going dialogue about any topic that the user cared to mention, without restriction, was that ELIZA did not use any domain knowledge about anything. Since ELIZA, most spoken dialogue systems have been closed-domain systems.

Voice interfaces for search engines have been developed. A speech interface to Google is described by Franz and Milch (2002). The big problem here is speech recognition, as the query is extremely short (median two words), and the vocabulary is large (over 100,000 words). This problem occurs in WikiTalk only when starting a totally new topic unrelated to previous topics. In that case, WikiTalk invites the user to spell the topic. During the conversation, topic shifts to related topics can be handled smoothly because WikiTalk extracts a small list of likely topics based on the NewInfo links. The speech recognizer only needs to recognize a vocabulary of 10 or 20 phrases (including the latest NewInfos and the system commands).

More recently, open-domain question-answering (QA) systems have appeared (Greenwood, 2008). These QA systems use question classifiers, search engines, ontologies, and answer extraction techniques. However, the basic aim of a QA system is to give the correct answer to a specific question, for example, "Q: What is the capital of Lorraine? A: Metz." QA systems are not normally intended to be conversational companions.

A Wikipedia-based question-answering system is described by Buscaldi and Rosso (2006). This QA system has a question type taxonomy and uses Wikipedia "category" entries (for example Category:Fruit) as a kind of ontology. The main aim is to use Wikipedia for validation of the answers, not as a source of topics for conversation.

The Ritel system (Rosset et al., 2006) combines an open-domain question-answering system with a spoken dialogue system. This allows the QA system to be more interactive, to ask clarification questions about the user's question. Kirschner (2007) describes different approaches to Interactive Question Answering. Follow-up questions in interactive QA systems are explored by Kirschner and Bernardi (2009).

These recent more interactive developments bring QA systems closer to dialogue systems. Nevertheless, the aim is still to find the answer to the question, not to talk about a topic. So far, QA systems do not suggest what would be an interesting question to ask.

7 Multimodal capabilities

Following previous work with a robot simulator (Jokinen and Wilcock, 2012a), we have implemented WikiTalk on the Aldebaran NAO robot (Jokinen and Wilcock, 2012b; Csapo et al., 2012). Using a real robot instead of a simulator has enabled us to include multimodal communication features for the robot, especially face-tracking and gesturing. These have

been integrated with the spoken conversation system. The robot needs to know whether the human is interested or not in the topic, and the human's proximity and gaze are important for this. Face-tracking is used to provide gaze information which is integrated in the interaction management. The robot also combines suitable nodding, gestures and body language with its own speech turns during the conversation.

For example, beat gestures are small hand movements that do not change the content of the accompanying speech but rather serve a pragmatic function, and emphasise and give rhythm to the speech. Beat gestures usually occur with NewInfos, serving a similar role as intonation to distinguish new and not expected information from the old and expected Topic information. In this way the communication is managed multimodally, and the visual management by gestures emphasises the least known elements to the partner so that the partner surely will notice and understand the new information.

To ensure maximal impact, the agents must make NewInfos as clearly available for the partner as possible, by using suitable lexical items, prosody (pitch, stress, volume, speed), and non-verbal means (gestures, gazing, face expressions), while the partner must be aware of these means in order to integrate the intended meaning in the shared context. Important topics in interaction management are thus related to information presentation: planning and generation of appropriate responses, giving feedback, and managing topic shifts.

An important factor in developing systems that can talk about interesting topics is assessing the level of interest of the user. There are two sides to this: first, how to detect whether the human conversational partner is interested in the topic or not, and second, what should the system do based on this feedback. The approaches to detecting the level of interest are part of the system's external interface, and the decisions about what to do based on this feedback are part of the system's internal management strategy. The external interface must clearly not be limited to purely verbal feedback, but must include intonation, eye-gaze, gestures, body language and other factors in order to assess the interest level correctly. The internal strategy for reacting appropriately to this feedback must decide what to do if the user is clearly interested, or is clearly not interested, and how to continue when the user's interest level is unclear.

Information about the evaluation of the Nao robot system based on the recorded user testing sessions at the 8th International Summer Workshop on Multimodal Interfaces, Metz, 2012 is given in (Csapo et al., 2012) and (Anastasiou et al., 2013).

Acknowledgments

The author is profoundly indebted to Kristiina Jokinen, whose work on the role of NewInfos has been an inspiration for two decades. Many thanks to Dimitra Anastasiou, Adam Csapo, Emer Gilmartin, Jonathan Grizou, JingGuang Han, and Raveesh Meena for implementing and evaluating WikiTalk and the multimodal interaction capabilities on the Nao robot at the 8th International Summer Workshop on Multimodal Interfaces, Metz, 2012.

References

Anastasiou, D., Jokinen, K., and Wilcock, G. (2013). Evaluation of WikiTalk - user studies of human-robot interaction. Proceedings of 15th International Conference on Human-Computer Interaction (HCI 2013).

- Buscaldi, D. and Rosso, P. (2006). Mining knowledge from Wikipedia for the question answering task. In *Proceedings of 5th Language Resources and Evaluation Conference (LREC 2006)*, Genoa.
- Csapo, A., Gilmartin, E., Grizou, J., Han, J., Meena, R., Anastasiou, D., Jokinen, K., and Wilcock, G. (2012). Multimodal conversational interaction with a humanoid robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2012)*, Kosice.
- Franz, A. and Milch, B. (2002). Searching the web by voice. In *Proceedings of 19th International Conference on Computational Linguistics (COLING 2002)*, pages 1213–1217, Taipei.
- Greenwood, M., editor (2008). *2nd Information Retrieval for Question Answering Workshop (IR4QA'08)*, Manchester. Workshop at COLING 2008.
- Jokinen, K. (2009). *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. John Wiley & Sons.
- Jokinen, K., Tanaka, H., and Yokoo, A. (1998). Context management with topics for spoken dialogue systems. In *Proceedings of Joint International Conference of Computational Linguistics and the Association for Computational Linguistics (COLING-ACL98)*, pages 631–637, Montreal, Canada.
- Jokinen, K. and Wilcock, G. (2011). Emergent verbal behaviour in human-robot interaction. In *Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011)*, Budapest.
- Jokinen, K. and Wilcock, G. (2012a). Constructive interaction for talking about interesting topics. In *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2010)*, Istanbul.
- Jokinen, K. and Wilcock, G. (2012b). Multimodal open-domain conversations with the Nao robot. In *Fourth International Workshop on Spoken Dialogue Systems (IWSDS 2012)*, Paris.
- Kirschner, M. (2007). Applying a focus tree model of dialogue context to interactive question answering. In *Proceedings of ESSLLI'07 Student Session*, Dublin, Ireland.
- Kirschner, M. and Bernardi, R. (2009). Exploring topic continuation follow-up questions using machine learning. In *Proceedings of NAACL HLT 2009: Student Research Workshop*, Boulder, Colorado.
- McCoy, K. F. and Cheng, J. (1991). Focus of attention: Constraining what can be said next. In C.L. Paris, W. S. and Mann, W., editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pages 103–124. Kluwer Academic Publishers.
- Rosset, S., Galibert, O., Illouz, G., and Max, A. (2006). Integrating spoken dialogue and question answering: The RITEL project. In *Proceedings of InterSpeech '06*, Pittsburgh.
- Weizenbaum, J. (1966). Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45.

