Double Topic Shifts in Open Domain Conversations: Natural Language Interface for a Wikipedia-based Robot Application

Kristiina Jokinen
University of Tartu, Estonia
University of Helsinki, Finland
kristiina.jokinen@helsinki.fi

Graham Wilcock
CDM Interact, Finland
University of Helsinki, Finland
gw@cdminteract.com

Abstract

The paper describes topic shifting in dialogues with a robot that provides information from Wikipedia. The work focuses on a *double topical construction of dialogue coherence* which refers to discourse coherence on two levels: the evolution of dialogue topics via the interaction between the user and the robot system, and the creation of discourse topics via the content of the Wikipedia article itself. The user selects topics that are of interest to her, and the system builds a list of potential topics, anticipated to be the next topic, by the links in the article and by the keywords extracted from the article. The described system deals with Wikipedia articles, but could easily be adapted to other digital information providing systems.

1 Introduction

Smooth human-like interactions to access data in knowledge bases have long been possible with advanced interaction technology. Current applications combine speech and language technology in mobile applications such as Siri, Cortana, and Alexa, to allow chat-like conversations related to topics that are of interest to the user, and the aim is to equip such interactive agents with more knowledge for enabling information exchange with wider topic domains and richer semantics. As argued by McTear et al. (2016), there is a big need for speech-based conversational interfaces that allow easy-to-use, natural, affective, and adaptable interactions with the user, while Jokinen (2009) pointed out that speech makes interactions more human-like and thus increases expectations about the system's competence in natural interaction. One of the bottlenecks for interactive systems has been the amount of data needed for successful interaction management, and usually systems have dealt with limited domains (bus timetables, flight information, pizza ordering, etc.) where the type and amount of knowledge allows the dialogues to be manually structured for the purposes of the task and the intentions of the user.

Robots and virtual agents have made conversational AI agents common and useful for various tasks where interaction with the user is needed. Their human-like appearance calls for more human-like communication, and research has focused on social interaction, multimodal issues, and affective computing for companion applications. Recently, chat systems or non-goal-oriented dialogue systems have also received much attention as they seem to encourage entertaining and engaging interactions as well as provide useful research platforms for studying emotion, social communication, and shared context. For instance, Zhou et al. (2016) describe evaluation of user engagement in a chatbot system, while Otsuka et al. (2016) study responses based on discourse relations.

While earlier work on dialogue systems was based on small, manually structured knowledge bases, we can now take advantage of very large internet-based resources and recent advances in machine learning to train robust interactive information-providing systems. Larger knowledge bases also open up possibilities for systems that are no longer restricted to one particular domain, and enable a move towards

Proceedings of the Open Knowledge Base and Question Answering (OKBQA) Workshop,
pages 59–66, Osaka, Japan, December 11 2016.

open-domain conversational systems which use the knowledge sources to provide chat-type interactions with no dialogue task other than being sufficiently entertaining. However, such open-domain conversational systems are challenging since meaningful interaction also needs to be addressed in terms of dialogue coherence. As argued in Jokinen (2009), conversations are not just a collection of separate questions and suitable answers, but form a conversational thread which exhibits the speaker's intentions of what they want to talk about (topics) as well as coherent and cooperative interaction management. Using the terminology of Grosz and Sidner's (1986) seminal model of discourse structure, dialogue coherence should be formulated in terms of the user's intention, attention, and linguistic processing levels.

Our work has focussed on building an interactive robot agent which converses with the user on the basis of the information found on internet web pages. The WikiTalk application (Wilcock, 2012; Jokinen and Wilcock, 2014) looks at conversational activity from the point of view of constructing a shared context in which the interlocutors exchange messages about interesting topics. The interlocutors' activities concern their reaction to the partner's presentation of new information, and various conversational management strategies which aim to catch the partner's attention, to build mutual understanding, and to keep the flow of information going. We hypothesise that this can be best done via a *double topical construction of dialogue coherence* which refers to discourse coherence taken into consideration on two levels; the evolution of dialogue topic via the interaction between the user and the robot system, and the creation of discourse topics via the content of the digital information article itself.

In this paper, we investigate mechanisms for topic introduction in conversational interactions with a humanoid robot, and discuss models for the computational management of the use of the web as the source of information through which a robotic agent can draw its "knowledge" for the interactions. The work extends open-domain dialogue management towards creating dialogues from any web content, but it also focusses the system on a particular goal in the same way as task-oriented dialogues: here the goal is to provide useful information based on existing web content and to help users to navigate and find the most interesting web pages with topics relevant to their individual interests. The paper is structured as follows. We discuss background for our work and present the problem in Section 2, then continue with the system overview in Section 3, discussion in Section 4, and finally conclude in Section 5.

2 Wikipedia and dialogue topics

In dialogue system design, one of the important issues is to equip the system with appropriate and sufficient domain information. This determines the type of questions the user can ask and the details of information that the system can talk about. The information needed for dialogue systems often already exists in some form, usually as a website, and the question is how to use this information for creating dialogues. In this paper, the first steps are described for transforming information automatically from websites into a natural language dialogue which can be used in building a robot dialogue system.

The focus is on how to present information in a manner that allows the user to follow the presentation and allows the system to anticipate the questions that the user may ask. It is important to notice that although our goal is to build an open-domain spoken dialogue system, we do not aim at a QA-type system that answers questions but rather at a chat-type dialogue system that can follow the user's topic shifts. Open-domain QA systems, such as IBM's Watson (Ferrucci, 2012), use sophisticated machine-learning techniques, question classifiers, search engines, ontologies, summarization, and answer extraction to enable efficient and accurate responses, but the aim of the system is still to find the correct answer to the question, not to hold a conversation about the topic as such. Interaction development has brought QA systems closer to dialogue systems, e.g. the RITEL system (Rosset et al., 2006) has a QA component which is used to ask clarification questions. However, QA systems are still intended to function primarily as interactive interfaces to information retrieval tasks rather than as conversational companions (see Moriceau et al. (2009) for an overview of information retrieval and automatic summarization systems).

In the WikiTalk application (Jokinen and Wilcock, 2014) the user can have a dialogue with the robot in which the robot talks fluently about an unlimited range of topics using information from Wikipedia. The system does not have typical task goals (book a hotel, get timetable information etc.), and is not a typical QA system that provides answers to particular questions. Rather, it aims to function on a more general conversational level and achieve the goal of "provide information on interesting topics" (as long as the user is interested in hearing about it, or the user can switch to a new interesting topic). It is thus important that the system can anticipate what the possible interesting continuations of the current topic

are, i.e. what kind of topical interests the Wikipedia article may bring forward. Related topics that the user may wish to continue the dialogue with are marked in Wikipedia with hyperlinks to other entries, so anticipated smooth topic shifts in conversations can be made to the relevant topics via the links.

Following the WikiTalk model, the user can query Wikipedia via the robot and have information from chosen articles read out by the robot. Wikipedia articles are considered as possible topics that the robot can talk about, while each link in an article is treated as new information that the user can shift their attention to, and ask for more information about. The paragraphs in the article are regarded as pieces of information that structure the main topic into subtopics, and they form the minimal units for presentation, i.e. a paragraph can be presented in one 'utterance' by the robot. A humanoid robot with movable arms and legs can also add non-verbal cues to enhance comprehension and to help the user to recognise the discourse level organisation of the text. We experimented with various gestures to provide structuring for the robot presentation, e.g. the robot uses gestures to emphasise the links while reading the text without recourse to explicit link menus, changes posture to mark turn-taking, and pauses after each paragraph to elicit feedback from the user whether to continue on the current topic or not.

In a spoken application, the anticipation of the topics that the user may want to know more about is important in order to assist the speech recognition component to arrive at the correct topical word. The existing method is simply to collect the links of the Wikipedia article and use these as the list of anticipated topics. Coherence of the interaction is thus based on the existing structure of the article and the links between the articles: they form the *first topical construction of dialogue coherence*.

However, the Wikipedia articles may also contain topics which are not currently linked to any other Wikipedia article, i.e. the wikification (Milne and Witten, 2008) of the author's text has not included these concepts in the set of linked concepts, or the article itself brings into the mind of the user topics which are triggered on the basis of the text but are not in the list of links. Our work in this paper addresses exactly this problem: how to anticipate suitable topics for the human-robot interaction when the topics are not explicitly marked in the wikification of the Wikipedia articles. Moreover, if the robot system also needs to relate to other digital resources than Wikipedia, e.g. digital news repositories or other webpages, then a more general anticipation method is necessary, since these resources may not have Wikipedia-type links available for smooth topic shift anticipation. The keyword extraction method to be described below is an alternative method to identify topics and forms the *second topical construction of dialogue coherence*. We call the two processes *double topical construction of dialogue coherence* as it includes two different but inter-related sets of topics to be created and managed by the system.

3 Topic anticipation

3.1 System overview

The WikiTalk system overview is given in Figure 1. The interaction with the user is handled by Conversation Manager which uses dialogue state representations to describe the current state of the conversation, and executes domain-independent dialogue tasks such as informing, requesting more information, clarifying speech input, and giving feedback.

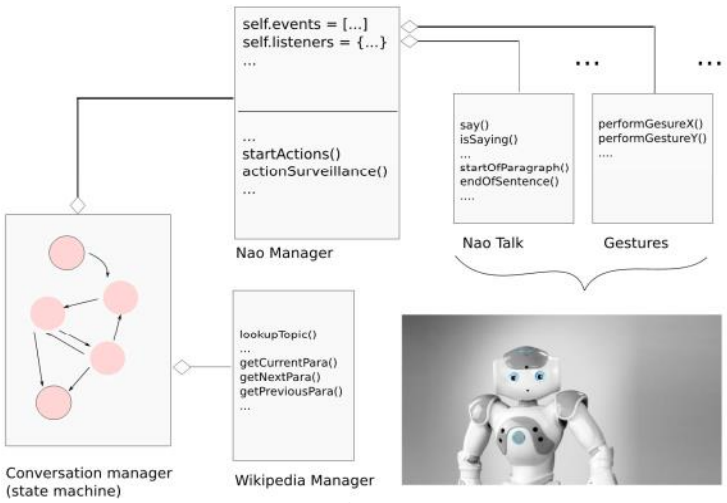

Figure 1 System overview (from Jokinen and Wilcock, 2014).

Conversation Manager receives possible topics and information about Wikipedia pages from Wikipedia Manager which takes care of the Wikipedia interface. The Nao Manager (called so because the robot platform is the Nao robot from Aldebaran) then renders the request via speech and gesture acts.

The domain-specific knowledge is provided through a topic tree (Jokinen et al., 1998; see Section 4), which suggests smooth topic shifts for the conversation. A topic tree is a structure built on the domain knowledge contained in the Wikipedia article. Domain-specific topics are represented by the keywords which describe the content of the topical article, and they form sub-nodes of the current topic node.

The system thus receives two types of possible topics: the links provided in the Wikipedia article itself as well as the keywords produced on the basis of the domain knowledge. Coherence of interaction management can thus rely on the new information provided either by the logic of the wikification of the concepts in the topical article, or by the information content of the article represented by keywords. In this manner interaction management is extended to cover various types of open-domain topics that Wikipedia gives rise to, automatically and without manual annotation, and the interaction between the robot and the user can be made topically richer and more natural.

3.2 Topic anticipation

The topic tree is built by the *Topic Anticipation module* which is part of Wikipedia Manager. It selects the keywords from the Wikipedia article using standard keyword extraction techniques. Figure 2 presents an overview of the system following Jokinen and Mikulas (2016), where the algorithm is described in more detail.

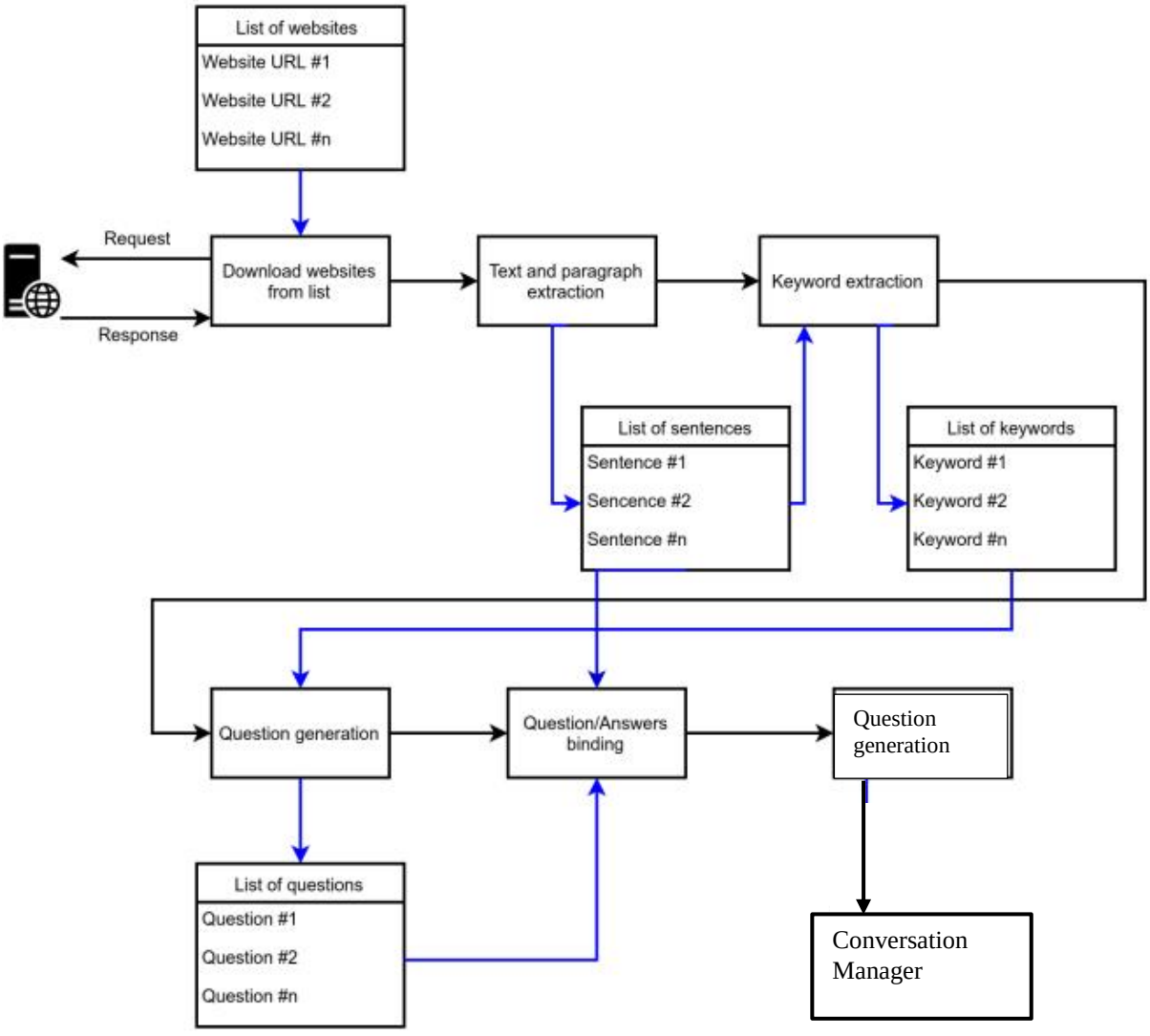

Figure 2 An overview of the topic anticipation module (following Jokinen and Mikulas 2016).

The algorithm first selects relevant sections from the web pages, then extracts text paragraphs from the sections, and cleans up the text before keyword extraction. The extracted keywords are used in question generation, i.e. suitable questions are generated based on the keywords.

Given a webpage like that in Figure 3 about Shakespeare, keywords are extracted from the paragraphs, and they represent a simple estimation of the page content. The keywords are determined via a Naïve Bayesian Filter, following Mooney (2005) and Matsuoka (2003). The frequency calculations currently use text corpora from Project Gutenberg, but we also plan to use larger corpora such as British National Corpus and Google Book Ngram Corpus (Lin et al. 2012) to get more balanced scores with respect to the text genre. The ratio f_p/f_s i.e. relative frequency of a word in the sample data (f_s) and in each extracted paragraph (f_p) measures how frequently a word occurs in the paragraph relative to the normal sample. The best *n* words are selected as keywords and in our experiments, the value 4 seems to provide the best balance between accurate content representations while also being small enough a number for question generation. By varying *n*, it is possible to experiment with a small vs. large number of possible keywords, i.e. vary the range of possible topics available for a conversation. In order to optimise the results for dialogue interactions, pruning may be necessary to select appropriate alternatives that accord with the users' preferred questions, or machine learning may be used to learn the user preferences.

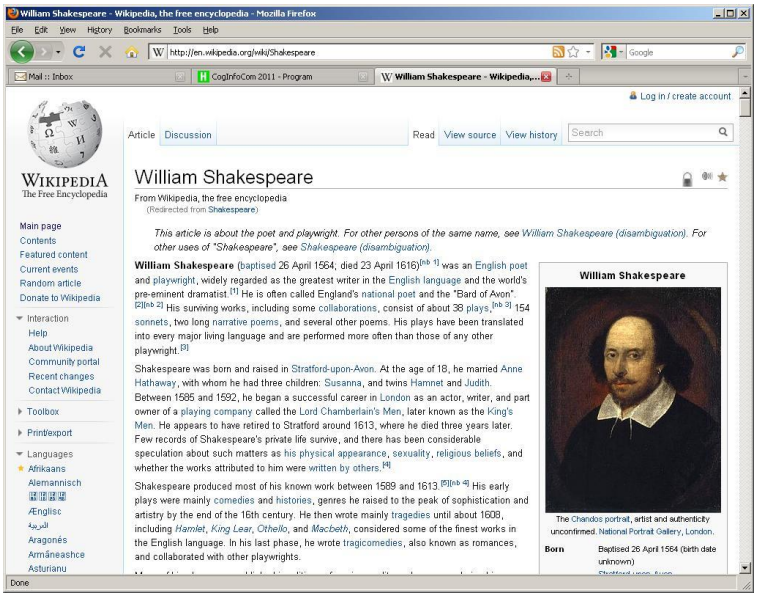

Figure 3 Wikipedia article about Shakespeare.

Some of the keywords extracted from the Wikipedia webpage about Shakespeare in Figure 3 are 'Stratford', 'Hathaway', 'London', 'tragedies', 'Macbeth', 'comedies', and 'romances', which are plausible next topics: places of residence, wife's name, and drama genres. Their scores are:

Stratford	2274900.0	Macbeth	1137450.0
Hathaway	1137450.0	romances	1137450.0
London	1137450.0	later	18957.5
tragedies	1137450.0	death	7109.1

In addition to anticipating the next topics in the conversation, the system's interaction strategies include anticipation of possible user questions to which appropriate answers should be provided. The extracted keywords can thus be used for generating questions which are expected to match the user's interest and include possible next topics that the user can request information about. For instance, in the above Shakespeare example, if the user wants to continue this topic, it is likely that she would like to know more about Stratford or Hathaway or Shakespeare's plays. The keywords thus function as links to Wikipedia articles which are suitable answers to such questions, either because of a direct match with an article title or via keywords related to the article content. The system described by Jokinen and Mikulas (2016) also has a list of template questions which can be used to produce questions by replacing the

placeholder with a keyword. A template *Tell me about X* would thus generate requests like *Tell me about Stratford* or *Tell me about Macbeth*. The templates can be used to help speech recognition to recognize anticipated topics, and if the user then utters such a request, the system can map the keyword to a relevant website.

3.3 Anticipating user interest in the extracted topics

An important property of a humanoid robot is its situatedness: the robot acts in the same space with the user, and this contributes to the immediate presence of the robot in the situation. Although robot reactions may be slow, this is similar to face-to-face human interactions where the listener can immediately give feedback on the presented information and the speaker can modify the presentation according to the listener's feedback. The speaker's anticipation of the partner's reaction as well as the listener's attendance to the speaker's presentation appear as co-creation of the discourse, and are manifested e.g. in producing feedback in the form of back-channelling and various non-verbal signals. We believe the Topic Anticipation module supports the above view of interaction: it operates both in the perception and generation phase, simultaneously as the robot tells the information to the user and observes their reaction to the presented information.

When developing systems that can talk about interesting topics with the user, a crucial factor is to assess the level of interest of the user. There are two sides to this: first, how to detect whether the partner is interested in the topic or not, and second, what should the system do based on this feedback The detection of the user's interest level belongs to the system's external dialogue interface and includes interpretation of the user's verbal and non-verbal feedback signals such as intonation, laughing, eye-gaze, nodding, and body movements, to assess her engagement in the interaction (see e.g. Jokinen and Wilcock (2012) and references therein). The decision about how to react to the various degrees of user engagement is part of the system's internal dialogue management, and how this is done is discussed e.g. in Jokinen & Wilcock (2014) and references therein. The interest level is specific to a particular topic, and may change in time. The user may show low interest in the current topic itself, but may show greater interest in a piece of new information that is mentioned.

4 Discussion

In dialogue management, topics are usually managed by a stack, which allows a convenient last-in-first-out mechanism to handle topics that have been recently talked about. However, the challenge in managing Wikipedia-based topics is how to convey the Wikipedia structure to the user so that the user can navigate within the new information links *and* refer to the content of the Wikipedia article she may want to know more about.

We use topic trees (cf. McCoy and Cheng 1990) in which topics are structured into a tree that enables more flexible management of the recent topics than a stack. *Topic* refers to the particular issue (Wikipedia article) that the speakers are talking about, and *NewInfo* is the part of the message that is new in the context of the current Topic (the paragraphs as the robot is reading the text, as well as the links in the article, and the extracted keywords).

Earlier research (as far as we know) has not been concerned with this kind of *double topical construction of dialogue coherence*, and we believe the work described in this paper is novel in that it tries to combine two topical structures: the development of human-robot interaction as coherent topic chains created through the interaction, and the recurring sentential topics that make the Wikipedia texts coherent as discourse. The computational management of the two topic structures and their development are taken care of by the two models: the "traditional" dialogue model is based on the user's interest in a particular topic and is responsible for driving the conversation forward with dialogue acts such as Question and Inform, while the discourse level possibility to create new topics through the lexico-referential topical progression is taken care of by the novel Topic Anticipation component performing keyword extraction. The meaningfulness of the whole interaction is thus built by anticipating possible topical questions via the links and via the extracted keywords, and then by the users' actually occurring choices of topics that they find interesting.

The dialogue coherence appears straightforward: we can rely on the link structure of Wikipedia to provide coherence for the dialogue, but also assume that the keyword extraction provides coherence for the possible continuation of the current topic to one of the keywords. It must be noted that Topic trees

created by the keywords and from user navigation via links from one Wikipedia page to another provide a different topic structure from the linguistically oriented topic structure formed from the sentences of the Wikipedia texts. For instance, in the above example of the Shakespeare text, the sentential subjects encode the recurrent topic (Shakespeare) of the paragraphs (subtopics) either directly or through lexical reiteration, superordination, meronymy, or co-reference. The point of departure chosen by the article's writer determines the discourse thematic position of these topics, and all other sentential topics are presented as hierarchically subordinate to it (e.g. *surviving works, Stratford-on-Avon, Anne Hathaway, early plays, tragedies*, etc.). The discourse topic in the webpage itself is constructed through the written coherent text, via lexico-referential topical progression, and it is a different process from the human-robot interaction concerning the robot telling the user about interesting topics.

5 Future work and conclusion

The paper has addressed issues related to *double topical construction of dialogue coherence* in the context of WikiTalk, an interactive robot interface to large digital information resources in the internet. The solution uses natural language processing methods to create automatically a list of possible topics for the robot to continue a coherent dialogue, on the basis of the webpage associated with the current topic of the conversation. The purpose of the work is to extend the system's current method, which uses the explicitly marked Wikipedia links to anticipate smooth topic shifts, with a new capability to anticipate topics which are not linked to another webpage, but which may still be interesting to the user based on the theme of the current webpage.

The work provides another viable avenue to integrate natural language interfaces to novel technological devices like robots, and uses the WikiTalk model to allow access to large digital resources in the internet. Compared with smartphones, tablets, smart watches, etc., a conversational robot interface features more *human* language properties which can be expected to make the query interface easier, more acceptable and accessible. For instance, autonomous robots can move and follow the user independently, rather than be carried in one's hand. This allows people who cannot hold or operate a small device in their hands to talk and hear about the topics they are interested in. Moreover, situated interactions enable multimodal communication which not only provides alternative ways to access the data, but encourages holistic communication between the user and the agent.

In many practical applications, new challenges appear for coordinating and managing online information with the help of natural conversation (e.g. teaching, meetings, non-goal-oriented conversations). Interaction with such applications requires dynamic tracking of dialogue topics and the user's focus of attention with respect to their interests and the actual situation. Thus models and techniques for tracking topics and focus of attention are important, and call for multidisciplinary approaches that combine interaction technology, AI-based system development, and communication studies.

Although the agent's communicative capability can become livelier and push natural language technology forward, the algorithms and methods still need further improvement and testing. For instance, the keyword selection could be elaborated and pruning of the keywords for final application be based on machine learning. Future work will also include more extensive evaluation with the robot agent. The current system has only been evaluated with respect to its operation and first impressions by the users, but a more systematic user study is scheduled to be conducted focussing on a system with keywords selected "as is" and keywords pruned for the purposes of interaction. The evaluation of the topic model in a practical application will also enable assessment of the effect of the humanoid robot's appearance on the user's experience and evaluation of the system, and whether the robot is able to capture the user's attention and contribute to their understanding and topic structuring by its own non-verbal signalling.

A demonstration of the robot interaction will be presented at the main conference (Wilcock et al, 2016) to substantiate the sketch of the dialogue interaction presented here.

Acknowledgements

We thank Mikulas Muron for his work on keyword extraction, and financial support from the Estonian Science Foundation project IUT 20-56 and the Academy of Finland grant n°289985.

References

Ferrucci, D.A. 2012. Introduction to "This is Watson". *IBM Journal of Research and Development*, vol. 56 no 3.4, pp. 1:1–1:15.

Foster, M. E. and Petrick, R. P. A. 2016. Separating representation, reasoning, and implementation for interaction management. In: Jokinen, K. and Wilcock, G. (Eds.) *Proceedings of the Seventh International Workshop on Spoken Dialogue System (IWSDS 2016),* Saariselkä, Finland.

Grosz, B.J. and Sidner, C.L. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3): 175-204.

Jokinen, K., Tanaka, H. and Yokoo, A. 1998. Context Management with Topics for Spoken Dialogue Systems. Proceedings of COLING 1998.

Jokinen, K. and Mikulas, M. 2016. Automated Questions for Chat Dialogues with a Student Office Virtual Agent. Proceedings of the Workshop on Chatbots and Conversational Agents (WOCCHAT), 16th International Conference on Intelligent Virtual Agents (IVA 2016), Los Angeles, U.S.A.

Jokinen, K. and Wilcock, G. 2012. Multimodal Signals and Holistic Interaction Structuring. *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012*). Mumbai, India.

Jokinen, K. and Wilcock, G. 2014. Multimodal open-domain conversations with the Nao robot. In: Mariani, J., Rosset, S., Garnier-Rizet, M., Devillers, L. (eds.) *Natural Interaction with Robots, Knowbots and Smartphones: Putting Spoken Dialogue Systems into Practice*, pp. 213–224. Springer.

Lin, Y., Michel, J-P., Lieberman Aiden, E., Orwant, J., Brockman, W. and Petrov, S. 2012. Syntactic Annotations for the Google Books Ngram Corpus. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.* Demo Papers, Jeju, Republic of Korea, Vol 2: 169–174.

Matsuoka, Y. 2003. Keywords extraction from single document using words co-occurrence statistical information. University of Tokyo. http://www.worldscientific.com/doi/abs/10.1142/S0218213004001466

McCoy, K. and Cheng, J. 1991. Focus of attention: Constraining what can be said next. In Paris, C.L., Swartout, W.R. and Moore, W.C. (Eds.) *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pp 103–124. Kluwer Academic Publishers.

McTear, M., Callejas, Z. and Griol, D. 2016. *The Conversational Interface*. Springer.

Milne, D. and Witten, I. H. 2008. Learning to link with Wikipedia. *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM),* pp. 509–518. New York, NY, USA.

Mooney, R.J. 2005. Text Mining with Information Extraction. Multilingualism and Electronic Language Management. University of Texas. http://www.cs.utexas.edu/~ml/papers/discotex-melm-03.pdf

Moriceau, V., San Juan, E., Tannier, A., and Bellot, P. 2009. Overview of the 2009 QA Track: Towards a Common Task for QA, Focused IR and Automatic Summarization Systems. *In Focused Retrieval and Evaluation, 8th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2009*. Brisbane, Australia, pages 355-365. Springer. Lecture Notes in Computer Science (LNCS 6203).

Otsuka, A., Hirano, T., Miyazaki, C., Higashinaka, R., Makino, T. and Matsuo, Y. 2016. Utterance Selection using Discourse Relation Filter for Chat-oriented Dialogue Systems. In: Jokinen, K. and Wilcock, G. (Eds.) *Proceedings of the Seventh International Workshop on Spoken Dialogue System (IWSDS 2016,)* Saariselkä, Finland.

Rosset, S., Galibert, O., Illouz, G. and Max, A. 2006. Integrating spoken dialogue and question answering: The RITEL project. *Proceedings of InterSpeech '06*, Pittsburgh.

Wilcock, G. 2012. WikiTalk: A spoken Wikipedia-based open-domain knowledge access system. *Proceedings of the COLING 2012 Workshop on Question Answering for Complex Domains*. pp. 57–69. Mumbai, India.

Wilcock, G., Jokinen, K., and Yamamoto, S. 2016. What topic do you want to hear about? A bilingual talking robot using English and Japanese Wikipedias. *Proceedings of COLING 2016*, Osaka, Japan.

Yu, Z., Xu, Z., Black, A.W, and Rudnicky, A. 2016. Chatbot evaluation and database expansion via crowdsourcing. In Proceedings of the RE-WOCHAT workshop of LREC, Portoroz, Slovenia.