Constructive Interaction for Talking about Interesting Topics

Kristiina Jokinen and Graham Wilcock

University of Tartu, Estonia and University of Helsinki, Finland
kristiina.jokinen@helsinki.fi, graham.wilcock@helsinki.fi

Abstract

The paper discusses mechanisms for topic management in conversations, concentrating on interactions where the interlocutors react to each other's presentation of new information and construct a shared context in which to exchange information about interesting topics. This is illustrated with a robot simulator that can talk about unrestricted (open-domain) topics that the human interlocutor shows interest in. Wikipedia is used as the source of information from which the robotic agent draws its world knowledge.

Keywords: open-domain dialogues, human-robot interaction, Wikipedia

1. Introduction

Natural language is used to exchange information, and the effective transfer of information is often taken as the main criterion for the success of interaction. Especially in the context of automatic services, the delivery of reliable and relevant information is an important goal for the design of such systems. Recently, however, one of the challenges for designing interactive systems has been identified as being related to social aspects of interactions: how to engage the partner in the interaction and keep their interest up so that the speaker can either deliver the message they intend to deliver, or can provide rapport and affection so as to create a mutual bond and an understanding relationship.

Engagement is a complex process that involves various multimodal cues and signals, and interaction research has focussed on the function and correlation of such multimodal communicative means as overlapping speech, gaze, facial displays, hand gestures, head movement, and body posture. For instance, Campbell and Scherer (2010) and Jokinen (2011) describe utterance density as a measurement for engagement. On the other hand, research with Embodied Conversational Agents (ECAs) has especially brought forward several types of behaviors that are important when conducting natural conversations between humans, and which are also necessary when supplying natural intuitive communication models for interactions between humans and ECAs (André and Pelachaud, 2010).

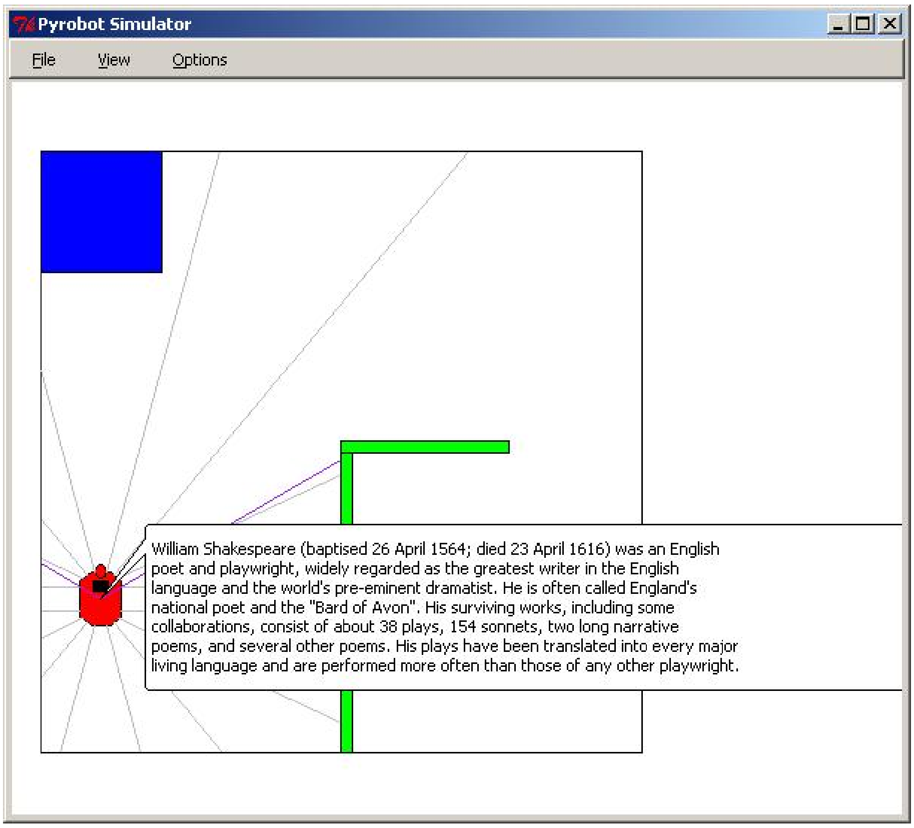
One of the important aspects in verbal and non-verbal communication is the actual topic that the speakers converse about, which may be either interesting or less attractive to the partner to discuss. Moreover, the presentation of information is pertinent to the success of communication: we know that the partners attend to the semantic content and the proposition of the message, and focus their attention on particular words that catch their attention and interest. The ability to keep the conversation going is important in many human-human social situations, and chatty conversational agents are also built for the Loebner prize competition. The competition is based on the Turing Test, where the interactive agents are to converse with the human user on any topic, and do it so well that for the human judge it is difficult to distinguish whether the partner is a real human or a computer agent.

Also in many practical applications interaction technology has to address challenges that concern engagement of the user in the interaction. The system has to coordinate the interaction and manage online information so that the pieces of new information that it intends to convey to the partner, can be used as a basis for natural conversation rather than a monologue on a particular topic. For instance, in teaching and learning situations, meetings and negotiations, such conversational capability is a useful skill. Furthermore, the users need to interact with other humans, and with intelligent robotic applications, and this requires dynamic tracking of dialogue topics and the users' focus of attention with respect to their interest and the actual situation. Models and techniques for tracking topics and focus of attention in interactive situations are thus important, and we aim to tackle the challenges using a multidisciplinary approach that combines interaction technology, AI-based systems, and communication studies.

This paper deals with speakers' interaction management strategies that are used to catch the partner's attention, to build mutual understanding, and to keep the flow of information going. We investigate mechanisms for topic management in conversational interactions, and concentrate on conversational activity where the interlocutors react to each other's presentation of new information and construct a shared context in which to exchange information about interesting topics. We demonstrate a robotic simulator that can talk about open-domain topics that are interesting to the human interlocutor. We use Wikipedia as the source of information from which the robotic agent draws its world knowledge.

2. Previous work

This paper continues previous work. Jokinen and Wilcock (2011) describe emergent verbal behaviour that arises when speech components are added to a robotics simulator. In the unmodified simulator the robot performs its activities silently. When speech synthesis is added, the first level of emergent verbal behaviour is that the robot produces spoken monologues giving a stream of simple explanations of its movements. When speech recognition is added, human-robot interaction can be initiated by the human, using voice commands to direct the robot's movements. In addition,



William Shakespeare (baptised 26 April 1564; died 23 April 1616) was an English poet and playwright, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. He is often called England's national poet and the "Bard of Avon". His surviving works, including some collaborations, consist of about 38 plays, 154 sonnets, two long narrative poems, and several other poems. His plays have been translated into every major living language and are performed more often than those of any other playwright.

Figure 1: Starting first topic: Shakespeare.

cooperative verbal behaviour emerges when the robot modifies its own verbal behaviour in response to being asked by the human to talk less or more.

The robotics framework supports different behavioural paradigms, including finite state machines, reinforcement learning, fuzzy decisions, neural networks and evolutionary algorithms. By combining finite state machines with the speech interface, spoken dialogue systems based on state transitions can be implemented for classical closed-domain form-filling dialogues such as flight reservations. These closed-domain dialogue systems exemplify emergent verbal behaviour that is robot-initiated: the robot asks appropriate questions in order to achieve the dialogue goal. A demo of these different levels of emergent verbal behaviour is described by Wilcock and Jokinen (2011).

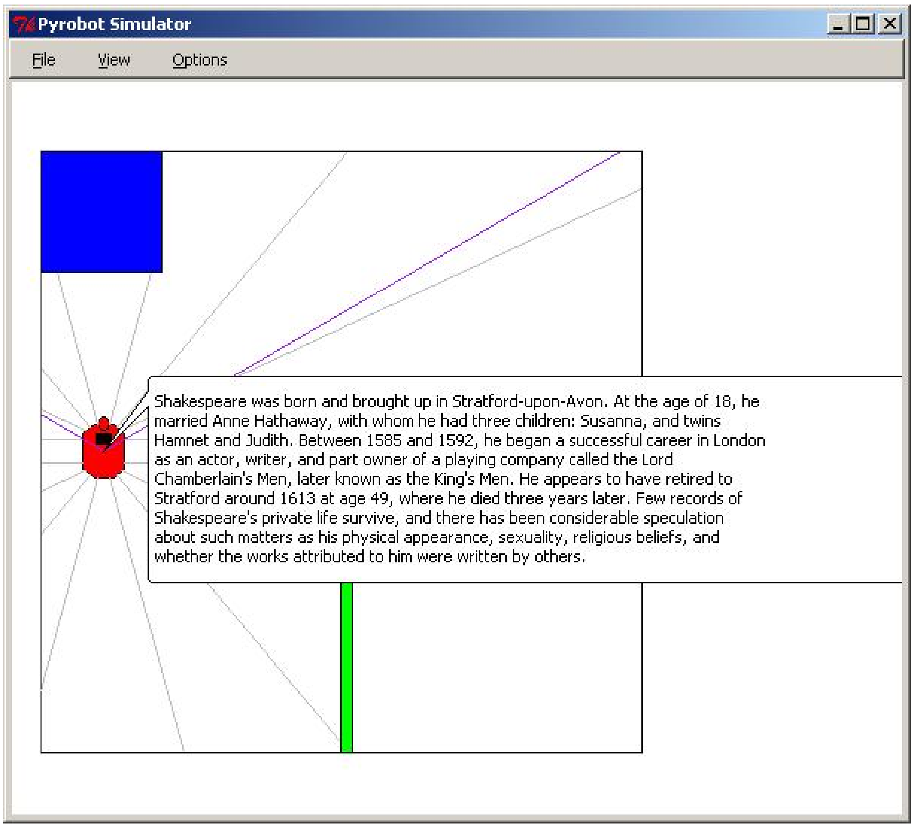
The next level of emergent verbal behaviour is open-domain conversational dialogues. Jokinen and Wilcock (2011) propose extending the robot's capabilities by using Wikipedia as a source of world knowledge. By exploiting ready-made paragraphs and sentences from Wikipedia, a robot can talk about a very wide range of open-domain topics. The example in Section 5. shows how the robot can change topics according to the human's interests. Although the specific topic will typically be human-initiated, as in the example, Wikipedia can also be used as a source of suggestions for new robot-initiated topics.

3. Constructive Dialogue Modelling

The theoretical basis of the interaction is drawn from the Constructive Dialogue Model (Jokinen, 2009), in which interaction management is regarded as coordinated action by rational agents. This AI-based approach integrates topic management, information flow, and the construction of shared knowledge in the conversation by communicative agents, and feedback is considered as the basic communicative obligation of the agents in order to build a common ground: building shared understanding of what has been exchanged in the conversation. The agents are engaged in activities whereby they exchange new information on a shared goal, and their communicative behaviour is based on their observations about the world as well as on their reasoning, within the dialogue context, about the effect of the exchanged new information on the underlying goals.

The new information is exchanged in the dialogue contributions. The speakers construct a shared context in which to resolve the underlying task, and their actions are constrained by communicative obligations arising from the particular activity they are engaged in and have a certain role in. The success of the interaction depends on the cognitive and emotional impact of the response on the hearer, and attention is paid to the planning and generation of appropriate responses, giving feedback, and topic management.

This model can be applied 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. The agents construct a shared



Shakespeare was born and brought up in Stratford-upon-Avon. At the age of 18, he married Anne Hathaway, with whom he had three children: Susanna, and twins Hamnet and Judith. Between 1585 and 1592, he began a successful career in London as an actor, writer, and part owner of a playing company called the Lord Chamberlain's Men, later known as the King's Men. He appears to have retired to Stratford around 1613 at age 49, where he died three years later. Few records of Shakespeare's private life survive, and there has been considerable speculation about such matters as his physical appearance, sexuality, religious beliefs, and whether the works attributed to him were written by others.

Figure 2: Continuing first topic: Shakespeare.

context in which the underlying goals can be achieved, and success of interaction depends on the cognitive and emotional impact of the action on the hearer.

To ensure maximal impact, the agents must make new information 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.

The main challenge lies in the grounding of language: constructing a shared knowledge of what the conversation is about and updating one's own knowledge accordingly. The focus of the research is on verbal and non-verbal signals that regulate the flow of information. Such aspects as looking at the conversational partner or looking away provide indirect cues of the partner's willingness to continue interaction, while gesturing can tell the partner that the item is important new information that the partner should focus their attention on. In this paper we focus especially on the agent's ability to engage the partner in the interaction by providing interesting information and interesting topics for discussion.

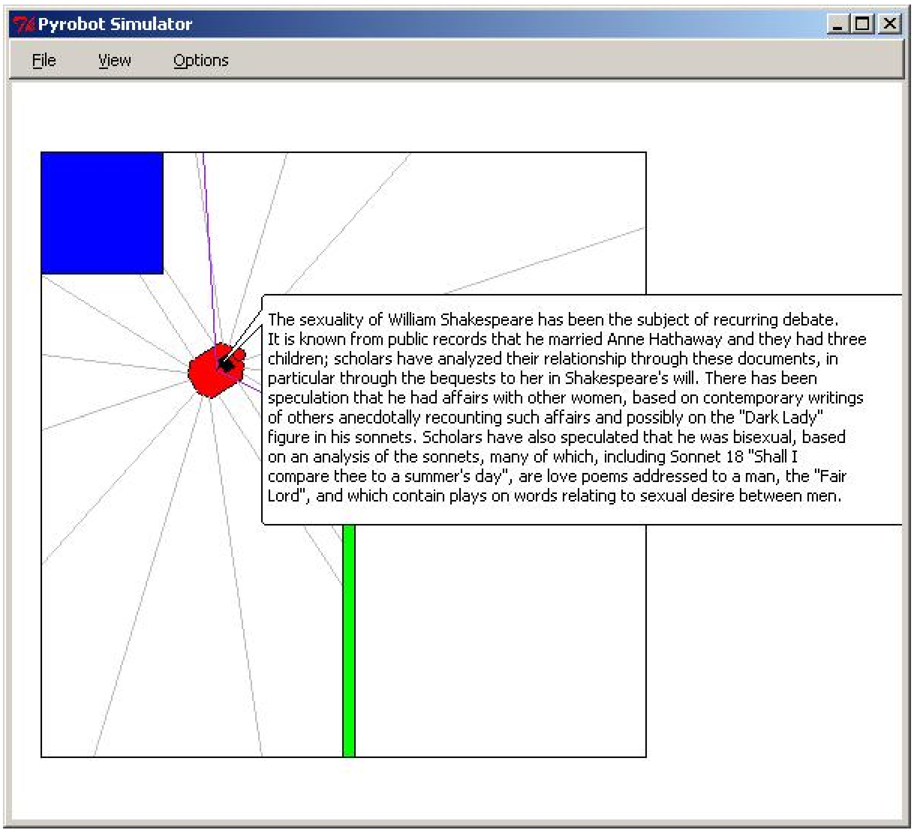
4. Topic trees and the Web

The organization of knowledge has always been one of the big questions. We can look for help with this question from the internet, in fact we can assume that world knowledge is somehow stored in the internet and we wish to take advantage of this.

The organization of knowledge into related topics is often done with the help of topic trees. Originally "focus trees" were proposed by McCoy and Cheng (1991) to trace foci in natural language generation systems. The branches of the tree describe 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 not very likely to occur, and if they do, they should be appropriately marked. The focus tree is a subgraph of the world knowledge, built in the course of the discourse on the basis of the utterances that have occurred so far. The tree both constrains and enables prediction of what is likely to be talked about next, and thus provides a top-down approach to dialogue coherence. The topic (focus) is a means to describe thematically coherent discourse structure, and its use has been mainly supported by arguments regarding anaphora resolution and processing effort.

Previously, topic trees were hand-coded which of course is time-consuming and subjective, or automatic clustering programs were used which have not been entirely satisfactory. Our approach to topic trees exploits the organisation of domain knowledge in terms of topic types found in the web, and more specifically in Wikipedia.

We use topic information in predicting the likely content of the next utterance, and thus we are more interested in the topic types that describe the information conveyed by utterances than the actual topic entity. Consequently, instead of



The sexuality of William Shakespeare has been the subject of recurring debate. It is known from public records that he married Anne Hathaway and they had three children; scholars have analyzed their relationship through these documents, in particular through the bequests to her in Shakespeare's will. There has been speculation that he had affairs with other women, based on contemporary writings of others anecdotally recounting such affairs and possibly on the "Dark Lady" figure in his sonnets. Scholars have also speculated that he was bisexual, based on an analysis of the sonnets, many of which, including Sonnet 18 "Shall I compare thee to a summer's day", are love poems addressed to a man, the "Fair Lord", and which contain plays on words relating to sexual desire between men.

Figure 3: Changing to a second topic: Shakespeare's sexuality.

tracing salient entities in the dialogue and providing heuristics for different shifts of attention, we seek a formalisation of the information structure of utterances in terms of the new information that is exchanged.

Wikipedia provides an extensive, freely available, open-domain and constantly growing knowledge source. We therefore use Wikipedia to produce robot contributions in open-domain dialogues, continuing the previous work described in (Jokinen and Wilcock, 2011).

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

5. Topic chains: an example

A concrete example of topic chains using Wikipedia is shown in Figures 1 to 6, using the robot simulator described in (Wilcock and Jokinen, 2011). The robot simulator can produce speech bubbles as well as actual speech, enabling these illustrative screenshots to be produced.

This specific sequence is initiated by the human asking the robot to talk about Shakespeare. The robot downloads the Wikipedia article about Shakespeare from the internet and starts reading the first paragraph as shown in Figure 1.

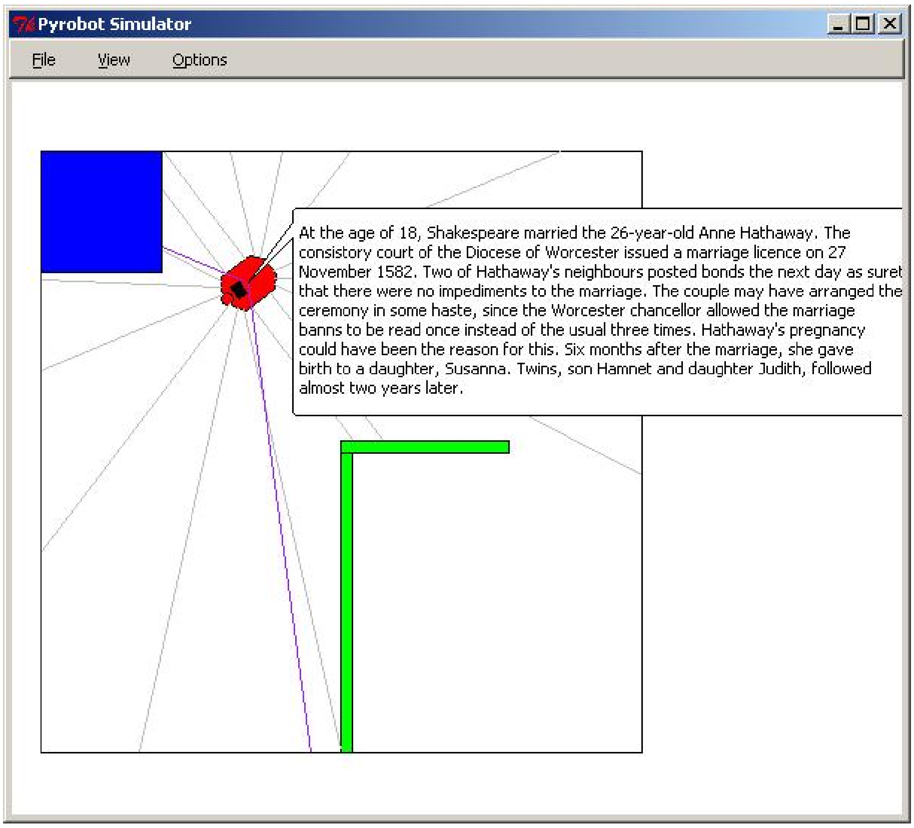
The most appropriate chunk size for a robot's spoken contribution will vary from a short phrase to a full sentence or a whole paragraph. This depends on several factors including the individual user (adult, child, expert, etc.), the communicative environment (quiet, noisy, etc.), and the previous interaction history (smooth, difficult, etc.). In these examples the chunk size is a whole paragraph, which may be suitable for a robot conversation with an adult in a quiet environment with a smooth previous interaction history.

In the example sequence, when the robot has spoken the first paragraph as shown in Figure 1, the human gives some feedback that is evaluated by the robot as sufficiently encouraging to continue the topic, so the robot decides to continue with the next paragraph as shown in Figure 2.

Shakespeare's physical appearance, sexuality and religious beliefs are all introduced as items of new information (NewInfos) near the end of the paragraph in Figure 2. As the human is relatively likely to ask about one of these NewInfos if any of them arouse sufficient interest, they are added to the list of specific items that the speech recognizer will listen for and be specifically ready to recognize.

Notice that the robot does not explicitly ask the user if they want to continue with one of these topics. It is assumed that the user will select the next topic as a natural continuation in the dialogue i.e. since the human is engaged in the dialogue.

In the example sequence, after the robot's contribution shown in Figure 2, the human asks about Shakespeare's sexuality. This is immediately recognized as one of the NewInfos that was likely to be selected for a topic shift. The



At the age of 18, Shakespeare married the 26-year-old Anne Hathaway. The consistory court of the Diocese of Worcester issued a marriage licence on 27 November 1582. Two of Hathaway's neighbours posted bonds the next day as suret that there were no impediments to the marriage. The couple may have arranged the ceremony in some haste, since the Worcester chancellor allowed the marriage banns to be read once instead of the usual three times. Hathaway's pregnancy could have been the reason for this. Six months after the marriage, she gave birth to a daughter, Susanna. Twins, son Hamnet and daughter Judith, followed almost two years later.

Figure 4: Continuing second topic: Shakespeare's sexuality.

robot therefore does not continue reading the main article about Shakespeare, but instead downloads a new Wikipedia article about Shakespeare's sexuality and starts reading the first chunk as shown in Figure 3.

The topic has now shifted to Shakespeare's sexuality. The human next gives feedback that is evaluated as showing sufficient interest in the new topic, so the robot continues with the next paragraph about Shakespeare's sexuality as shown in Figure 4.

Shakespeare's children Susanna, Hamnet and Judith are all introduced as NewInfos near the end of this paragraph. These NewInfos are therefore added to the list of items that the speech recognizer will be ready to recognize.

Now the human asks about Shakespeare's son Hamnet and this is recognized as one of the NewInfos that was likely to be selected for a topic shift. The robot therefore quits the article about Shakespeare's sexuality, and downloads another Wikipedia article about Shakespeare's son Hamnet and starts reading the first chunk as shown in Figure 5.

Next the human asks about "Hamlet" which is mentioned in the text about Hamnet and is recognized as one of the NewInfos likely to be selected for a topic shift. The robot therefore stops talking about Hamnet, downloads a new Wikipedia article about Hamlet (Shakespeare's play) and starts reading the first chunk as shown in Figure 6.

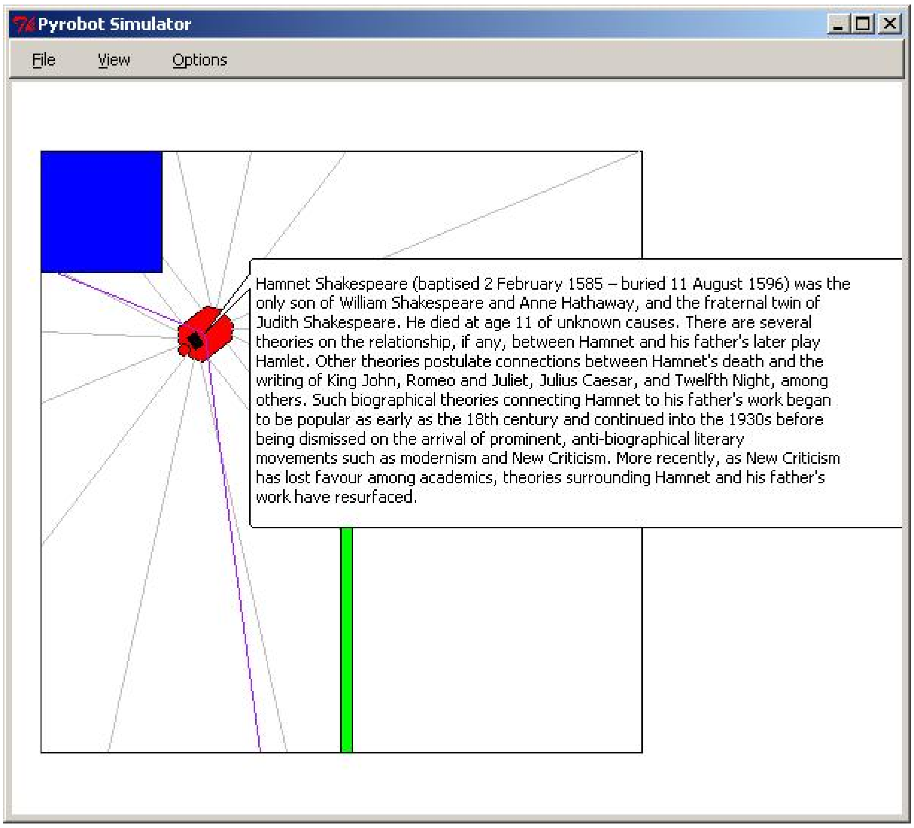
6. Assessing the level of interest

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.

In future work (see Section 7.), we aim to use a real robot instead of a simulator. This will enable us to include multimodal communication features for the robot, especially gaze-tracking and gesturing. These need to be 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 gaze is important for this. Eye-tracking equipment will be used to provide gaze information so that the role of gaze-tracking can be integrated in the interaction management. The robot should also combine suitable gestures and body language with its own speech turns during the conversation. This requires a model of when to gesture and what kind of gestures to use (hands, head, body).

Note that the interest level is specific to a particular topic, including potential new topics. The user may show low interest in the current topic itself, but may show greater in-



Hamnet Shakespeare (baptised 2 February 1585 – buried 11 August 1596) was the only son of William Shakespeare and Anne Hathaway, and the fraternal twin of Judith Shakespeare. He died at age 11 of unknown causes. There are several theories on the relationship, if any, between Hamnet and his father's later play Hamlet. Other theories postulate connections between Hamnet's death and the writing of King John, Romeo and Juliet, Julius Caesar, and Twelfth Night, among others. Such biographical theories connecting Hamnet to his father's work began to be popular as early as the 18th century and continued into the 1930s before being dismissed on the arrival of prominent, anti-biographical literary movements such as modernism and New Criticism. More recently, as New Criticism has lost favour among academics, theories surrounding Hamnet and his father's work have resurfaced.

Figure 5: Changing to third topic: Shakespeare's son Hamnet.

terest in a piece of new information that is mentioned. This feedback is then used by the system, which switches topics smoothly to the new information of greater interest.

7. Future work

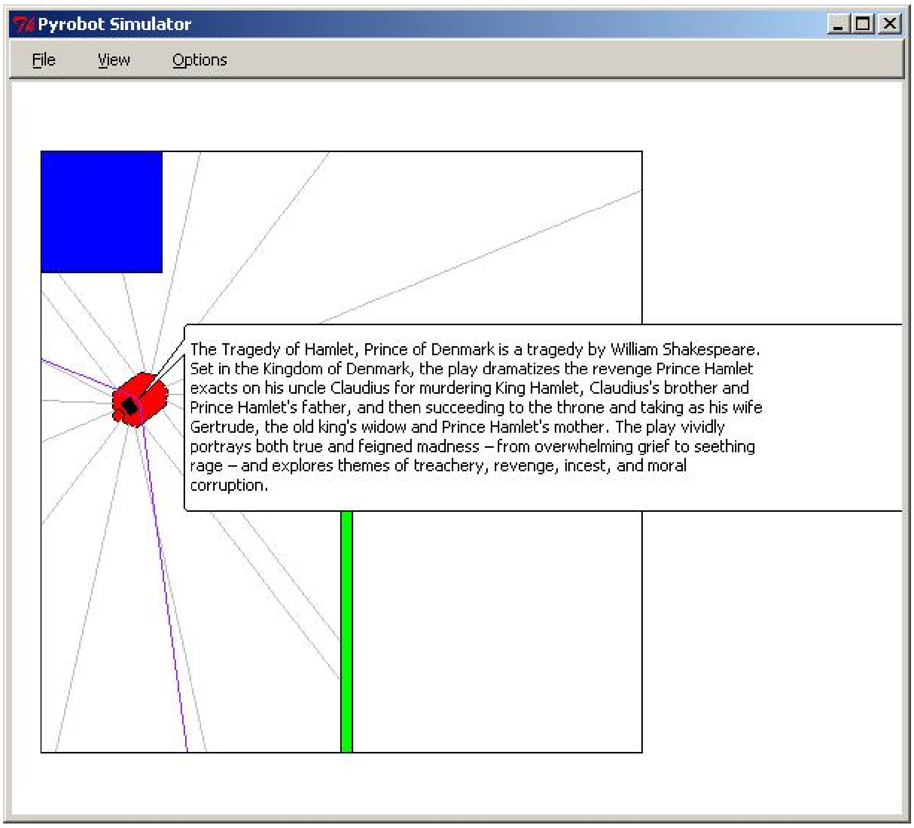
As future work, we plan to export the dialogue system from the Pyrobot simulator to a real robot. We then expect to be able to integrate speech and gesture so as to support the presentation of information.

For example, beat gestures are small hand movements that do not change the content of the accompanying speech but rather they serve a pragmatic function, and emphasise and give rhythm to the speech. Often beats are synchronized with the spoken emphasis, i.e. the stroke (the most energetic part of the gesture) occurs at the intonationally most prominent syllable of the accompanying speech segment (Kendon, 2004).

Gestures can also mark discourse structure. The pragmatic gesture, or what Jokinen and Vanhasalo (2009) called "stand-up gesture", signals to the partner that something important is to come next, and thus they direct the listener's attention to what is going to be said. Moreover, beat gestures usually occur with the new information, i.e. they serve a similar role as the intonation to distinguish new and not expected information from the topic, or old and expected information. In this way, the communication is managed in a multimodal way and the visual management by gestures is to emphasise the least known elements to the partner so that the partner surely will notice and understand the new information. There will of course be other multimodal signals as well but the synchrony of gestures and intonation has a particular significance (Jokinen, 2010).

One notable signal is the manner and frequency of feedback in conversation (Misu et al., 2011): whether one provides frequent verbal and non-verbal feedback concerning the basic enabling aspects of communication (contact, perception and understanding), or whether one tends to assume that the enablements hold as long as the actual interaction takes place, and thus no backchannelling is necessary in an explicit manner. Also the way feedback particles are used to express evaluation and acceptance of the given information or how the partners fill in each others' utterances vary a lot. Much of the conversational information exchange relies on the assumptions and presuppositions that are not necessarily made explicit in the course of the interaction. The context is an important source of information, and one of the necessary conversational skills is to know how to enable the right type of contextual reasoning: the participants should observe each others' reactions and changes in emotional and cognitive states.

We aim to build models to match human topic tracking possibilities with the linguistic-pragmatic competence of the robot, and thus ultimately to develop practical interactive agents. As human-robot interactions get more common and also more complex, the models for interaction must be based on a better understanding of topic tracking and basic mechanisms of conversational strategies that are crucial for flexible and intuitive interaction management. We also plan to experiment with the robot to assess the naturalness of the interactions with respect to the user's engagement.



The Tragedy of Hamlet, Prince of Denmark is a tragedy by William Shakespeare. Set in the Kingdom of Denmark, the play dramatizes the revenge Prince Hamlet exacts on his uncle Claudius for murdering King Hamlet, Claudius's brother and Prince Hamlet's father, and then succeeding to the throne and taking as his wife Gertrude, the old king's widow and Prince Hamlet's mother. The play vividly portrays both true and feigned madness – from overwhelming grief to seething rage – and explores themes of treachery, revenge, incest, and moral corruption.

Figure 6: Changing to a fourth topic: Shakespeare's play Hamlet.

8. References

Elisabeth André and Catherine Pelachaud. 2010. Interacting with embodied conversational agents. In F. Cheng and K. Jokinen, editors, *Speech Technology: Theory and Applications*, pages 123–150. Springer.

Nick Campbell and Stefan Scherer. 2010. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In *Proceedings of 11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, Makuhari, Japan.

Kristiina Jokinen and Minna Vanhasalo. 2009. Stand-up gestures - annotation for communication management. In C. Navarretta, P. Paggio, J. Allwood, E. Ahlsén, and Y. Katagiri, editors, *Proceedings of the NoDaLiDa workshop on Multimodal Communication - from Human Behaviour to Computational Models. NEALT Proceedings Series, Vol. 6*, pages 15–20, Tartu.

Kristiina Jokinen and Graham Wilcock. 2011. Emergent verbal behaviour in human-robot interaction. In *Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011)*, Budapest.

Kristiina Jokinen. 2009. *Constructive Dialogue Modelling: Speech Interaction and Rational Agents*. John Wiley & Sons.

Kristiina Jokinen. 2010. Gestures and synchronous communication management. In A. Esposito, N. Campbell, C. Vogel, A. Hussein, and A. Nijholt, editors, *Development of Multimodal Interfaces: Active Listening and Synchrony*, pages 33–49. Springer.

Kristiina Jokinen. 2011. Turn taking, utterance density, and gaze patterns as cues to conversational activity. In *Proceedings of the International Conference on Multimodal Interaction (ICMI-2011) Workshop on Multimodal Corpora for Machine Learning (MMC)*, Alicante, Spain.

Adam Kendon. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

Kathleen F. McCoy and Jeannette Cheng. 1991. Focus of attention: Constraining what can be said next. In W.R. Swartout C.L. Paris and W.C. Mann, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics)*, pages 103–124. Kluwer Academic Publishers.

Teruhisu Misu, Etsuo Mizumaki, Yoshinori Shiga, Shinichi Kawamoto, Hisashi Kawai, and Satoshi Nakamura. 2011. Analysis on effects of text-to-speech and avatar agent on evoking users' spontaneous listener's reactions. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 77–89, Granada.

Graham Wilcock and Kristiina Jokinen. 2011. Adding speech to a robotics simulator. In *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pages 371–376, Granada.