Bridging Gaps Between Planning and Open-Domain Spoken Dialogues

Kristiina Jokinen

Abstract In social media, Wikipedia is the outstanding example of a collaborative wiki. After reviewing progress in open-domain question answering systems, the paper discusses a recent system, WikiTalk, that supports open-domain dialogues by using Wikipedia as a knowledge source. With the collaboratively-written sentences and paragraphs from Wikipedia, the WikiTalk system apparently succeeds in enabling "open-domain talking". In view of recent advances in web-based language processing, the paper proposes steps towards open-domain "listening" that combine anticipated progress in open vocabulary speech recognition with recent developments in named entity recognition, where Wikipedia is now used as a dynamically updated knowledge source instead of fixed gazetteer lists. The paper proposes that Wikipedia-based open-domain talking and open-domain listening will be combined in a new generation of web-based open-domain spoken dialogue systems. Technological and social development affects our interaction with the environment: interactive systems are embedded in our environment, information flow increases, and interaction becomes more complex. In order to address challenges of the complex environment, to respond to needs of various users, and to provide possibilities to test innovative interactive systems, it is important to investigate processes that underlie human-computer interaction, to provide models and concepts that enable us to experiment with various types of complex systems, and to design and build tools and prototypes that demonstrate the ideas and techniques in a working system. In this article, I will discuss the "gap" between dialogue management and response planning and focus on the communicatively adequate contributions that are produced in the context of a situated robot agent. The WikiTalk system supports open-domain conversations by using Wikipedia as the knowledge source, and a version of it is implemented on the Nao-robot.

K. Jokinen (✉)
University of Helsinki, Helsinki, Finland
e-mail: kristiina.jokinen@helsinki.fi

N. Gala et al. (eds.), *Language Production, Cognition, and the Lexicon*,
Text, Speech and Language Technology 48, DOI 10.1007/978-3-319-08043-7_20

Keywords Wikitalk interaction · Open-domain dialogues · Newinfo · Topic trees · Planning · Generation

1 Prologue

In the mid-1990's I organised an ECAI workshop *GAPS AND BRIDGES: New Directions in Planning and Natural Language Generation*, together with Michael Zock and Mark Maybury. It was my first international workshop, and I was excited at the possibility to collaborate with two famous senior researchers of the field.

The workshop focussed on the planning of communicatively adequate contributions, and especially on the gap which at that time was recognized between natural language generation and AI-based planning for autonomous cooperative systems. In NLG, a focus shift directed research from grammatical well-formedness conditions towards exploration of the communicative adequacy of linguistic forms, while dialogue system research investigated how natural and rational communication could be equipped with NLG techniques so as to be able present the message to the user in a flexible way.

The gap has since been bridged, or at least it seems less deep, thanks to research and development on interactive systems and response planning. Generation challenges, e.g. Challenge on Generating Instructions in Virtual Environments (GIVE), have brought the NLG tasks closer to the planning of communicative contributions, while spoken dialogue systems need a module which effectively corresponds to a NLG component to be able to produce system output (see e.g. Jokinen and Wilcock 2003, or Jokinen and McTear 2009 for an overview). A recent indication of the mutual interests is a shared session which has been planned to take place in the SIGDial (Special Interest Group in Disourse and Dialogue) and the INLG (International Conference on Natural Language Generation) conferences organised by the respective communities in the summer 2014, in order to "highlight the areas of research were the interests of the two communities intersect and to foster interaction in these area".

The workshop topics seem timely and relevant even today, after 20 years of the original workshop, although they need to be formulated in a slightly different way. In fact, it is possible to argue that new bridges are needed to overpass the gaps between intelligent agents and open-domain generation tasks. On one hand, research in speech-based dialogue systems has extended communicative adequacy to cover not only grammatical forms of written language, but also meaningful exchanges of "ungrammatical" spoken utterances. An established view in dialogue research is that speaking is a means for achieving communicative goals, and dialogues are jointly constructed by the partners through communicatively appropriate utterances which can overlap with each other time-wise, and consist of elliptical structures as well as of discourse particles and backchannelling elements. Thus interactive agents, ranging from speech-enabled applications to situated

conversational robot companions, need to reason on the communicative context, including previous dialogue, physical situation, and the partner's knowledge and interest, in order to interpret the utterances and to engage the partner in the conversation. In generation research, on the other hand, information retrieval and summarization techniques have allowed NLG to extend research toward question-answering systems and thus the context also plays an important role in the planning and realization of responses: it has impact on the interpretation of the question and the information relevant to the answer, as well as on the user's knowledge, preferences, and interest regarding the topic of the question.

In this article, I will discuss the "gap" between dialogue management and response planning and focus on the communicatively adequate contributions that are produced in the context of a situated robot agent. The WikiTalk system supports open-domain conversations by using Wikipedia as the knowledge source (Wilcock and Jokinen 2012, 2013; Jokinen and Wilcock 2013), and a version of it is implemented on the Nao-robot (Csapo et al. 2012). The article is structured as follows. Section 2 reviews recent progress in open-domain interactive systems. Section 3 presents the WikiTalk application and discusses our approach to an open-domain dialogue system which can talk about any topics found in Wikipedia. Section 4 addresses the dichotomy of topic and New information, and Sect. 5 addresses the issues concerning Topic trees. Some notes on generation are presented in Sect. 6, and conclusions and future prospects are discussed in Sect. 7.

2 Open-Domain Interactive Systems

Open domain spoken dialogue systems that aim at serving as conversational companions must be capable of talking about any topic that the user introduces. The WikiTalk system (Wilcock 2012; Jokinen and Wilcock 2012) proposes to meet this requirement by using Wikipedia as a knowledge source. Wikipedia is a collaboratively produced encyclopaedia and it is constantly updated, so the range of topics that the WikiTalk system can talk about is unrestricted, and continuously growing. Contrary to traditional dialogue systems, dialogue management in WikiTalk is not based on a task but on the user's interest in the dialogue topics and on the system's ability to engage the user in an interesting conversation. The interaction management thus resembles the Question-under-Discussion approach (Ginzburg 1996), implemented as the Information State Update model in TrindiKit (Traum and Larsson 2003): the structure of the dialogues is determined by the information flow, and the dialogue is managed by updating the system's information state according to the user's questions and introduction of relevant pieces of information in the dialogue context. An important difference, however, is the availability of the topics in WikiTalk. In TrindiKit, the QUDs are limited to relevant information in a particular task, while in WikiTalk, topics are open to any information for which an article can be found in Wikipedia.

The most famous open-domain dialogue system was, and still is, ELIZA (Weizenbaum 1966). ELIZA could maintain an on-going dialogue, with no restriction on the topics that the user might care to mention. However, this was possible precisely because ELIZA did not use any domain knowledge about anything. Moreover, the user soon noticed that the dialogues lacked a goal and coherence: the system could maintain the dialogue for only a few turns and there was no global coherence or structure in the replies.

Modern versions of Eliza use chatbot technology based on the Artificial Intelligence Markup Language (AIML) standard. Such applications are designed specifically for web-based interaction on mobile devices such as handsets and tablets. For instance, Alice (http://www.alicebot.org/) provides a chatbot personality with a human-like face and interactive features, and it can be used on website or mobile app to chat with the user. However, the limit of the Alice framework is that it requires a hand-tailored database on the basis of which interaction takes place. Although the chatbot applications may sound fairly free and natural, they still require manually built domain models and question-answer pairs for smooth operation.

On the other hand, research with Embodied Conversational Agents (ECAs) has especially brought forward multimodal interaction, focussing on the different types of multimodal signalling that are important in human-human natural conversations, and which are also necessary when supplying natural intuitive communication models for interactions between humans and ECAs (André and Pelachaud 2010; Misu et al. 2011).

Recently, a new type of question-answering (QA) systems have appeared (Greenwood 2006) that are open-domain in the sense that the user can ask a question about any topic. One of the most famous ones of the new QA systems is IBM's Watson (Ferrucci 2012), whereas Apple's SIRI exhibits personal assistant and knowledge navigator with a capability to answer questions which are not directly related to its knowledge-base. Open-domain QA systems use sophisticated machine-learning techniques, question classifiers, search engines, ontologies, summarization, and answer extraction techniques to enable efficient and accurate response. Moriceau et al. (2009) give an overview of the information retrieval and automatics summarization systems, and Franz and Milch (2002) discuss various issues related to voice enabled search. Evaluation of such complex systems is also complicated, and e.g. El Ayari and Grau (2009) provide a glass-box evaluation framework for QA systems.

Different approaches to Interactive Question Answering are reviewed by Kirschner (2007). Although these more interactive developments have brought QA systems closer to dialogue systems, the aim of a QA system is still to find the correct answer to the question, not to hold a conversation about the topic as such. For example, an interaction may consist of a question "What is the second largest city in France?" and of the answer "Marseille." Efforts have also been made to build more interactive QA systems by combining them with aspects of a spoken dialogue system. For example, in the RITEL system (Rosset et al. 2006) the QA component has a capability to ask clarification questions about the user's question.

The combination of RITEL with another QA-system, QAVAL, extends the system with different answer extraction strategies which are then merged at different levels to the final answer (Grappy et al. 2012). However, QA systems are still primarily intended to function as interactive interfaces to information retrieval tasks, rather than as conversational companions.

In this context, a notable exception is the WikiTalk system that can be described from two points of view: it is a QA system in that it operates as an "open-domain knowledge access system" (Wilcock 2012), and it is a conversational dialogue system in that it allows "talking about interesting topics" (Jokinen and Wilcock 2012, 2013; Wilcock 2012). WikiTalk uses Wikipedia as its knowledge source, and by dynamically accessing the web, WikiTalk differs from traditional QA systems in that it is able to maintain a conversation about the topic introduced by the user. Wikipedia has been used by question-answering systems, as described for example by Buscaldi and Rosso (2006). However, their main aim is to use Wikipedia for validation of answers, not as a knowledge source for conversations: Wikipedia "category" entries are used as a kind of ontology which the QA's question type taxonomy can base its answers. The application domain envisaged in WikiTalk is not a fancy chatbot that provides clever answers on a predefined domain, but rather an interactive "agent" which has its cognitive capability extended by internet knowledge.

3 The WikiTalk Application

The WikiTalk (Jokinen and Wilcock 2012, 2013; Wilcock 2012) is an interactive application that allows the user to query and navigate among Wikipedia articles. By using Wikipedia as its knowledge source, WikiTalk is an open-domain spoken dialogue system as compared with traditional task-based dialogue systems, which operate on a closed-domain, finite application database.

The WikiTalk system works as a web application with a screen interface, but the implementation on the Nao humanoid robot greatly extends its natural dialogue capability. As described in Csapo et al. (2012), the robot implementation includes multimodal communication features, especially face tracking and gesturing. Face-tracking provides information about the user's interest in the current topic, while suitable gesturing enables the robot to emphasise and visualise its own information presentation. The human's proximity to the robot and their focus of visual attention are used to estimate whether the user follows the robot's presentation, whereas head nodding, hand gestures, and body posture are combined with the robot's own speech turns to make its presentations more natural and engaging. Figure 1 shows some users interacting with the Nao WikiTalk system during the ENTERFACE summer school 2011, and an annotated video of a Wikipedia-based open-domain human-robot dialogue can be seen at: http://vimeo.com/62148073.

The theoretical foundation of WikiTalk is Constructive Dialogue Modelling (CDM, Jokinen 2009), which integrates topic management, information flow, and

Fig. 1 Users interacting with the Nao WikiTalk

the construction of shared knowledge in the conversation by communicative agents. According to CDM, interlocutors are rational agents who coordinate and control their interaction in cooperation. Moreover, the agents monitor their partner's behaviour and give feedback to each other concerning the basic enablements of communication: Contact, Perception, Understanding, and Reaction (cf. Allwood 1976). Contact and Perception are understood as modelling the agent's awareness of the communication, while Understanding and Reaction concern the agent's intentional and cooperative behaviour: producing a semantic interpretation of the partner's utterance and to the planning and generation of one's own behaviour as a reaction to it, respectively. Signalling whether the basic enablements are fulfilled (the person hears what is said, understands the meaning of the partner's message, or is willing to be involved in the interaction) is often done via non-verbal and multimodal means, i.e. not explicitly by words but by head movements, facial expressions, gesturing, and body posture.

According to the CDM, dialogue management should support interaction that *affords* natural information flow (Jokinen 2009). In the context of WikiTalk, the main challenge is to present Wikipedia information in a way that makes the structure of the articles clear. The users should easily navigate among the topics that interest them, be able to pick up links for new information, and select new topics. WikiTalk keeps track of what is currently salient in the interaction (a model of the interlocutor's attention), and anticipates what is the likely next topic (a model of the interlocutor's communicative intentions). An interactive WikiTalk also distinguishes between two conditions: the user shows interest and allows the system to continue on the current topic, or the user is not interested in the topic and the system should stop or find some other topic to talk about. The interaction model thus includes a user model and a representation for the partner's mental states, to keep track of the topics being talked about and the user's interest and attitude towards the presented information.

The conversational strategy of the WikiTalk agent is designed to be verbose, with a goal of initiating topics which are likely to engage the user in the conversation. The dialogue control model in WikiTalk uses a finite-state approach, and Fig. 2 (next page) shows a pertinent state transition diagram (this diagram also shows speech recognition states, cf. Wilcock 2012). The diagram differs from traditional finite state models in that dialogue states are related to the information flow ("select New Topic", "continue Topic", etc.), not to specific domain-related

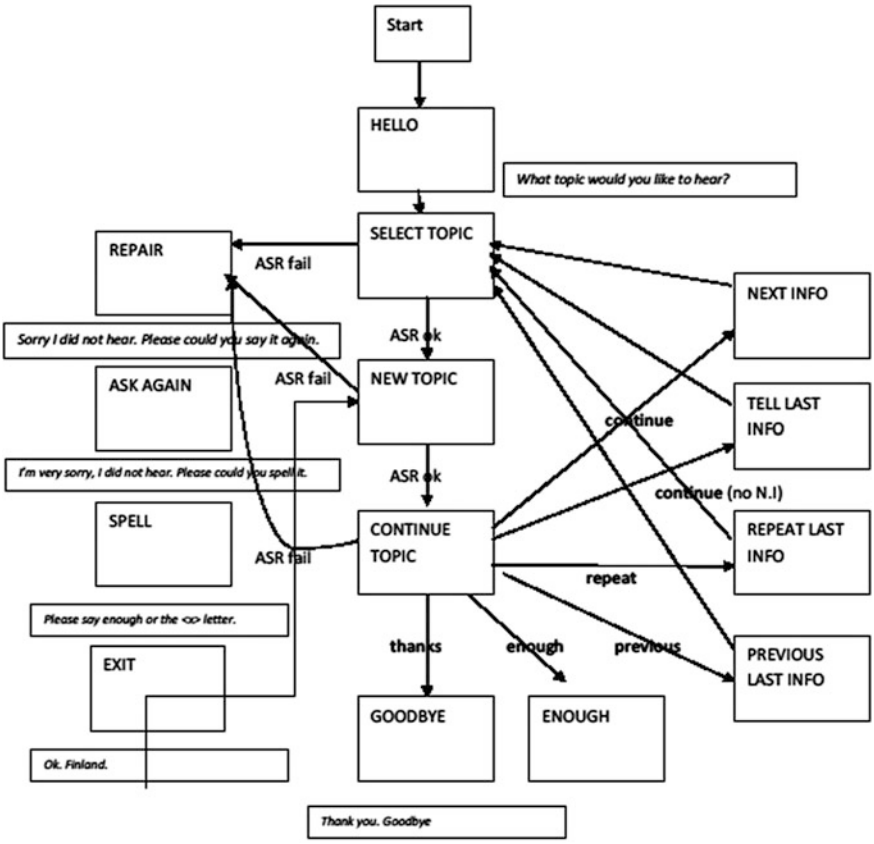

Fig. 2 A state transition diagram for WikiTalk

knowledge states (such as "know departure day" or "know destination city"). The state transitions concern the exchange of information with the user, and they are managed with the help of techniques related to topic-tracking and topic shifting. The dialogue actions are thus reduced to very general actions, namely presenting information and listening to the partner's response, rather than being related to particular task-domain actions. This approach makes it possible for a finite number of states to manage dialogues with an infinite number of topics.

When a new topic is selected, the WikiTalk system gets the text for the topic from Wikipedia and divides it into chunks (paragraphs and sentences) suitable for spoken dialogue contributions. The system then manages the presentation of the chunks according to the user's reaction. The user can respond verbally or non-verbally, and WikiTalk thus needs to be able to "listen" and "see", i.e. it needs to understand the user's verbally expressed commands (such as "continue"), and also interpret the user's multimodal behaviour (such as looking away to signal one is not interested). If the user shows interest in the

topic, by explicitly asking for more information or multimodally signalling their curiosity on the topic, the WikiTalk system continues with presenting the next chunk. At the end of each chunk, the user has an opportunity to continue with the next chunk, to ask for the same chunk to be repeated, or to go back to the previous chunk about the current topic. The user also can initiate a new topic, i.e. shift the topic to another one.

To manage user-initiated topic shifts, WikiTalk follows the user's changing interests by using hyperlinks from Wikipedia articles. The user is likely to pick up an interesting concept in the presentation, and by using this as a keyword, explicitly ask for more information about the particular topic. WikiTalk assumes that the topic shift is usually to a link introduced in the article (cf. hypertext navigation), and it can thus anticipate the course of the conversation by treating all the wiki-article's linked concepts as expected utterance topics that the user can pick up as the next interesting topic. For instance, if WikiTalk provides information about Marseille and says "Further out in the Bay of Marseille is the Frioul archipelago", the user can say "Frioul archipelago?" and WikiTalk will smoothly switch topics and start talking about the Frioul archipelago. From the dialogue management point of view, the system always has a set of concepts that it expects the user to pick up as smooth continuations of the current topic, assuming that the user is interested in the topic. In order to continue with such a smooth topic-shift, the user just says the name of the interesting NewInfo.

The user may also introduce a brand new topic, in which case the WikiTalk agent establishes the topic as a new selected topic. It may or may not be relevant to the previous topic, but its "newness" value for the system comes from the fact that it is not in the expected topic list but outside the current expectations. If the user does not initiate any topics, the system tries to engage the user by suggesting new topics. This is done by system checking Wikipedia for interesting topics in the daily "Did you know?" or "On this day" sections, and suggesting randomly some of these topics for the user. Again, if the user is interested in hearing more from the particular topic, the user is likely to indicate their interest by explicitly requesting "tell me more", or implicitly inviting WikiTalk to talk more about the issue with the help of multimodal signalling, e.g. raising eyebrows or uttering "really" with a raising intonation. At the moment, WikiTalk understands user utterances consisting of short commands or simple keywords, but we are experimenting with complex natural language utterances. Depending on the speech recognition engine, this works fairly well for English. The same applies for different speakers and accents: the ASR expects the user to speak fairly standard form of language.

It is also possible that the user wants to interrupt the current chunk without listening to it all, and ask to skip forward to the next chunk on the same topic. If WikiTalk is interrupted, it stops talking and explicitly acknowledges the interruption. It then waits for the user's input, which can range from telling the systems to continue, to go back to an earlier chunk, to skip forward to the next chunk, or to switch to a new topic.

4 NewInfos and Topics in WikiTalk

CDM follows the common grounding models of dialogue information (Clark and Brennan 1991; Traum 1994) in that the status of information available in a particular dialogue situation depends on its integration in the shared context. New information needs to be grounded, i.e. established as part of the common ground by the partner's acknowledgement, whereas old information is already grounded, or part of the general world knowledge. Normally the knowledge of a dialogue system is included in the system's task model that contains concepts for the entities, events and relations relevant for the task in hand, and the concepts are instantiated and grounded in the course of the dialogue.

In WikiTalk, however, the system's knowledge consists of the whole Wikipedia. Wikipedia articles are regarded as possible *Topics* that the robot can talk about, i.e. elements of its knowledge. Each link in the article is treated as a potential *NewInfo* to which the user can shift their attention by asking for more information about it. If the user follows a link in the article, the linked article thus becomes a new Topic and the links in it will be potential NewInfos. The conceptual space in WikiTalk is thus quite unlike that in a closed-domain task models: it consists of a dynamically changing set of Wikipedia articles and is structured into Topic and NewInfos according to the article titles and the hyperlinks in the articles.

The paragraphs and sentences in the article are considered propositional chunks, or pieces of information that form the minimal units for presentation. To distinguish the information status of these units of presentation, we use the term *focus text*: this is the paragraph that WikiTalk is currently reading or presenting to the user and which is thus at its focus of attention. We could, of course, apply the division of Topic and NewInfo also to paragraphs if we consider the Wikipedia articles only in terms of being presented to the user. However, we wanted to have an analogous structure for the Wikipedia article reading as we already had used for sentential information, where the concepts of Topic and NewInfo are instantiated as concepts that can be talked about and which form a concept space for the surface generator. In WikiTalk, Topic and NewInfo refer to the particular issues that the speakers can talk about or decide to talk about next (Wikipedia articles and the hyperlinks therein which form a conceptual network to navigate in), while the parts of the message (paragraphs) that are new in the context of the current Topic, are called focus texts. The focus texts provide new information about the current Topic (the title of the Wiki-article), but they cannot be selected by the user directly as something the user wants to talk about (e.g. it is not possible to issue a command in WikiTalk to read the third paragraph of the article Marseille). The focus text paragraphs are only accessible by the system when it reads the topical article. Unlike articles and hyperlinks, the focus texts are not independent "concepts" or referents in the conceptual space that can be referred to, but more like closely and coherently related pieces of information associated with a specific topic (wiki-article).

It must be emphasized that dialogue coherence is considered straightforward in WikiTalk; discourse relations between consecutive utterances rely on the structure of Wikipedia. Since the articles have already been written as coherent texts and the links between the articles have been inserted so that they make the articles into coherent hypertexts, we can assume that by following the topics and the NewInfo links in Wiki-articles the listener is able to infer what the connection between the topics is. We make a strong assumption in that the user selects links to continue dialogue, rather than any other words in the Wiki-article. The users can, of course, select any article as their next topic, and this is often the case if the user explores the Wikipedia randomly. On the other hand, most users are already used to navigating through Wikipedia and using the hyperlinks to select the next topic, so in WikiTalk, they simply follow the same principle. However, one of the relevant questions in WikiTalk is how to point to the users which of the words are linked and which are not—in speech this is not as easy as in visual texts. In the robot implementation, Nao WikiTalk can use the whole repertoire of communicative means, i.e. it uses rhythmic gesturing to mark the linked words.

Situated dialogue systems impose requirements on the generation of multimodal responses, e.g. to build models that determine appropriate prosody and the appropriate type of hand gesturing to accompany a spoken utterance. We will not go into details of these, but refer to André and Pelachaud (2010) for multimodal aspects, and to the seminal work by Theune (2000) on marking of pitch accents and phrasal melodies for the realizer to synthesize the correct surface form, or to introductory textbook like Holmes and Holmes (2002).

5 Topic Trees and Smooth Topic Shifts

In dialogue management, topics are usually managed by a stack, which conveniently handles topics that have been recently talked about. However, stacks are a rather rigid means to describe the information flow in cases where the dialogues are more conversational and do not follow any particular task structure. We prefer topic trees, which enable more flexible management of the topics. The trees can be traversed in whatever order, while the distance of the jumps determines the manner of presentation of the information.

Originally "focus trees" were proposed by McCoy and Cheng (1991) to trace foci in NL 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. The tree both constrains and enables prediction of likely next topics, and provides a top-down approach to dialogue coherence.

The notion of a topic (focus) has been a means to describe thematically coherent discourse structure, and its use has been mainly supported by arguments

regarding anaphora resolution and processing effort. WikiTalk uses topic information in selecting likely content of the next utterance (the links contained in the article), and thus the topic tree consists of Wiki-article titles that describe the information conveyed by the utterance (the article that is being read). We can say that the topic type or theme is more important than the actual topic entities. The WikiTalk system will not do full syntactic parsing, but will identify chunk boundaries (paragraph and sentence endings), so instead of tracing salient discourse entities and providing heuristics for different shifts of attention with respect to these entities, WikiTalk seeks for a formalisation of the information structure in terms of Topic and NewInfo that deal with the article titles and links.

In previous research, the world knowledge underlying topic trees was hand-coded, and this of course was time-consuming and subjective. In WikiTalk, it is the Wikipedia which is the "world knowledge" of the system, and our topic trees are a way to organise domain knowledge in terms of topic types found in the web. The hypertext structure is analogous to the linking of world knowledge concepts (although not a graph but a network), and through the interaction with the user, the system selects topics and builds a topic tree. The topic shifts which occur following the information structure in the Wikipedia will be *smooth* topic shifts, while the shifts which the user introduces and which are not part of the immediate information structure are called *awkward*. Consequently, smooth topic shifts are straightforward continuations of the interaction, but awkward shifts require that WikiTalk marks the shift verbally. This maintains the interaction coherence and clear.

6 Notes on Generation

In the standard model of text generation systems (Reiter and Dale 2000), information structure is recognised as a major factor. This model usually has a pipeline architecture, in which one stage explicitly deals with discourse planning and another stage deals with referring expressions, ensuring that topic shifts and old and new information status are properly handled. As already mentioned, spoken dialogue systems impose further requirements on generation, such as how to handle prosody, but there are some fundamental issues involved too, stemming from the facts that in spoken interaction, the listener also immediately reacts to the information presented, and the speaker can modify the presentation online based on the listener's feedback. This kind of anticipation of the partner's reaction and immediate revision of one's own behaviour brings us to the old NLG question of *Where does generation start from?* Previously it has been argued that attempts to answer this question push the researchers on sliding down a slippery slope (McDonald 1993) in that the starting point seems to evade any definition. However, considering generation in interactive systems, we can argue that it starts simultaneously with interpretation, in the perception and understanding phase of the presented information. In other words, generation starts already when one is

listening to the partner, as a reaction to the presented information. Although the actual realisation of the thoughts as spoken language appears later on (and is regulated by turn-taking conventions), the listeners can also produce immediate feedback in the form of backchannelling and various non-verbal signals.

The model of generation in WikiTalk loosely follows that introduced by Jokinen et al. (1998). In this model, response planning starts from the information focus, *NewInfo*, which can be thought as the content of the speaker's intention. The generator's task is to convey this message to the partner, so it decides how to present NewInfo to the user: whether to realise just the "naked" NewInfo by itself, or whether to add appropriate hedging information that would help the partner to understand how the NewInfo is related to the joint goals of the dialogue. For this, the manager creates an *Agenda*, a set of specifically marked domain concepts which have been designated as relevant in the dialogue. The Agenda is available for the generator, which can freely use the concepts in the agenda in order to realise the system's intention, but is not forced to include all the concepts in its response.

The prototype response generator described by Jokinen and Wilcock (2001) and Jokinen and Wilcock (2003) has a simple pipeline including an aggregation stage, a combined lexicalization and referring expressions stage, and a surface realization stage. In WikiTalk, the generation does not deal with the sentence realisation, since the Wikipedia articles are read aloud as such (using a TTS-system), while the system utterances are simple questions and canned phrases. However, the notions of Agenda and NewInfo are used when the system deals with the "information chunks", i.e. with the paragraphs and sentences of the articles. Following the NewInfo-based model, the WikiTalk Agenda is used to keep track of the "concepts", i.e. the focus texts that belong to the current topical article, and the generator can then select from Agenda those texts that will be realised and read to the user. All focus text paragraphs are marked as NewInfo, i.e. as potential information to be conveyed to the user, and the system realises them by selecting one at the time to be read aloud to the user. As mentioned earlier, the NewInfo presentation continues depending on the user's reaction: the next focus text will be presented to the user if the user seems interested or explicitly requests "continue".

7 Conclusion

A communicatively competent interactive system should provide communicatively adequate responses. In the Gaps and Bridges workshop, this was addressed by inviting submissions e.g. on *interactions between situational, motivational (speaker and addressee goals), cognitive and linguistic constraints*; as well as on *the effect of the various constraints on the generation process as a whole (resource-bounded agency and planning constraints; open-world assumption; time and space constraints)*. In this article I have returned back to the pertinent issues presented in the workshop, and considered issues related to the two above mentioned workshop themes. In particular, I have discussed a situational robot

application WikiTalk and its dialogue management model that supports open-domain interactions. It is noticed that in spoken interactive systems, the generation of responses can be seen as starting already when the system is listening to the partner, or in other words, besides open-domain talking, it is necessary to have "open-domain listening". As is clear, there is much still to be done in order to bridge the gap and address the workshop themes, but much active research is being conducted, and rapid progress can be expected.

References

Allwood, J. (1976). *Linguistic communication as action and cooperation*. Gothenburg Monographs in Linguistics 2. University of Gothenburg.

André, E., & Pelachaud, C. (2010). Interacting with embodied conversational agents. In F. Cheng, & K. Jokinen (Eds.), *Speech technology: Theory and applications* (pp. 123–150). Berlin: Springer.

Buscaldi, D., & Rosso, P. (2006). Mining knowledge from Wikipedia for the question answering task. In *Proceedings of 5th Language Resources and Evaluation Conference* (LREC 2006), Genoa.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: APA Books.

Csapo, A., Gilmartin, E., Grizou, J., Han, J. G., Meena, R., & Anastasiou, D., et al. (2012). Multimodal conversational interaction with a humanoid robot. In *Proceedings of 3rd IEEE International Conference on Cognitive Infocommunications* (CogInfoCom 2012), Kosice.

El Ayari, S, & Grau, B. (2009). A framework of evaluation for question-answering systems. *ECIR, Lecture notes in computer science* (vol. 5478, pp. 744–748). Berlin: Springer.

Franz, A., & Milch, B. (2002). Searching the web by voice. In *Proceedings of 19th International Conference on Computational Linguistics* (COLING 2002) (pp. 1213–1217). Taipei.

Ferrucci, D. A. (2012). Introduction to this is watson. *IBM Journal of Research and Development, 56*(3.4), 1:1–1:15.

Ginzburg, J. (1996). Interrogatives: Questions, facts and dialogue. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 385–422). Blackwell: Blackwell Textbooks in Linguistics.

Grappy, A., Grau, B., & Rosset, S. (2012). Methods combination and ML-based re-ranking of multiple hypothesis for question-answering systems. In *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data* (pp. 87–96). Avignon, France, April 2012.

Greenwood, M. A. (2006). Open-domain question answering. PhD Thesis,Department of Computer Science, The University of Sheffield.

Holmes, J. N., & Holmes, W. J. (2002). *Speech synthesis and recognition*. UK: Taylor Francis Ltd.

Jokinen, K. (2009). *Constructive dialogue modelling: Speech interaction and rational agents*. New York: Wiley.

Jokinen, K., & McTear, M. (2009). *Spoken dialogue systems.Synthesis lectures on human language technologies*. San Rafael, CA: Morgan and Claypool.doi:10.2200/S00204ED1V01Y200910HLT005.

Jokinen, K., Tanaka, H., & Yokoo, A. (1998). Planning dialogue contributions with new information. *Proceedings of the ninth international workshop on natural language generation* (pp. 158–167). Ontario: Niagara-on-the-Lake.

Jokinen, K., & Wilcock, G. (2003). Adaptivity and response generation in a spoken dialogue system. In J. van Kuppevelt, & R. W. Smith (Eds.), *Current and new directions in discourse and dialogue*. (pp. 213–234). UK: Kluwer Academic Publishers.

Jokinen, K., & Wilcock, G. (2012). Constructive interaction for talking about interesting topics. In *Proceedings of Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul.

Jokinen, K., & Wilcock, G. (2013). Multimodal open-domain conversations with the nao robot. In J. Mariani, L. Devillers, M. Garnier-Rizet, & S. Rosset (Eds.), *Natural interaction with robots, knowbots and smartphones—putting spoken dialog systems into practice*. Berlin: Springer.

Kirschner, M. (2007). Applying a focus tree model of dialogue context to interactive question answering. In *Proceedings of ESSLLI'07 Student Session*, Dublin, Ireland.

McCoy K. F., & Cheng, J. (1991). Focus of attention: Constraining what can be said next. In C. Paris, W. Swartout, & W. Mann (Eds.), *Natural language generation in artificial intelligence and computational linguistics*, (pp. 103–124). UK: Kluwer Academic Publishers.

Misu, T., Mizumaki, E., Shiga, Y., Kawamoto, S., Kawai, H., & Nakamura, S. (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* (pp. 77–89), Granada.

Moriceau, V., SanJuan, E., Tannier, X., & 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, (pp. 355–365). Springer Verlag. Lecture Notes in Computer Science (LNCS 6203).

McDonald, D. (1993). Does natural language generation start from a specification?.In: H. Horacek & M. Zock (Eds.), New concepts in natural language generation(pp. 275–297). London: Pinter Publishers.

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

Reiter, E., & Dale, R. (2000). *Building natural language generation systems*.Cambridge: Cambridge University Press. Reissued in paperback in 2006.

Theune, M. (2000). From data to speech: language generation in context. Ph.D. thesis, Eindhoven University of Technology.

Traum, D. R. (1994) A computational theory of grounding in natural language conversation, TR 545 and Ph.D. Thesis, Computer Science Dept., U. Rochester, December 1994.

Traum, D., & Larsson, S. (2003). The information state approach to dialogue management. In J. van Kuppevelt and R. Smith (Eds.), *Current and new directions in discourse and dialogue* (pp. 325–353). South Holland: Kluwer.

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.

Jokinen, K., & Wilcock, G. (2001). Pipelines, templates and transformations: XML for natural language generation. In *Proceedings of the 1st NLP and XML Workshop* (pp. 1–8). Tokyo.

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

Wilcock, G. & Jokinen, K. (2013). Towards cloud-based speech interfaces for open-domain coginfocom systems. In *Proceedings of the 4th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) 2013*, Budapest, Hungary.

Wilcock, G., & Jokinen, K. (2011). Adding speech to a robotics simulator. In R. Lopez Delgado, et al. (Eds.) *Proceedings of the Third International Conference on Spoken Dialogue Systems: Ambient Intelligence*. (pp. 375–380). Granada, Spain: Springer Publishers.