Small Talk is more than Chit-Chat

Exploiting Structures of Casual Conversations for a Virtual Agent

Nikita Mattar and Ipke Wachsmuth

Artificial Intelligence Group, Bielefeld University
Universitätsstr. 25, 33615 Bielefeld, Germany
[nmattar|ipke]@techfak.uni-bielefeld.de

Abstract. An approach of improving the small talk capabilities of an existing virtual agent architecture is presented. Findings in virtual agent research revealed the need to pay attention to the sophisticated structures found in (human) casual conversations. In particular, existing dialogue act tag sets lack of tags adequately reflecting the subtle structures found in small talk. The approach presented here structures dialogues on two different levels. The micro level consists of meta information (speech functions) that dialogue acts can be tagged with. The macro level is concerned with ordering individual dialogue acts into sequences. The extended dialogue engine allows for a fine-grained selection of responses, enabling the agent to produce varied small talk sequences.

1 Introduction

Research in the field of Embodied Conversational Agents has shown that it is not sufficient to restrict conversations between agents and humans to task-oriented topics. Findings suggest that small talk supports the deepening of relationships between virtual agents and human interaction partners. Especially when dealing with interactions over the long run it is inevitable to enable the agent to develop a close relationship to the human interactant. This trend is reflected in the emergence of new research areas of "more sociable" agents like companion agents. For details cf. [3].

First attempts of integrating small talk into task-oriented dialogues were restricted to common topics like the weather. While this is sufficient to fill short gaps between tasks, recently the focus shifted towards more elaborate small talk capabilities in order to further enhance the bonding between agent and human.

In some cases, even small talk can be regarded as task-oriented. Since it serves the purpose of establishing relationships it might be a very important goal for people to successfully engage in small talk with others. While a lot of people might complain that they do not like to participate in small talk, others are not capable of doing it. Still these people might benefit from training small talk. To enable a virtual agent to coach a human interlocutor doing small talk, the agent has to have a clear representation of small talk. Furthermore, the agent should perform small talk on different levels of complexity and therefore must be able to select from different sequences and strategies for conducting small talk.

The paper is structured as follows. In the section to follow we discuss the nature of small talk and casual conversations in human dialogue and briefly review relevant research on this issue in the virtual agents domain. In Section 3, the main section, we introduce our approach on how to improve the small talk capabilities of a virtual agent architecture by tagging dialogue acts with meta information in order to achieve a variety of small talk sequences. In Section 4 we describe how we plan to evaluate our system, and present our ideas how this work is to be continued.

2 Related Work

2.1 Small Talk and Casual Conversations

Small talk has been defined as a "conversation about things which are not important, often between people who do not know each other well" [1]. According to Schneider [16] small talk can be classified as a special kind of casual conversation (which is influenced by social distance), in that small talk is more likely to happen if the social distance is greater, whereas casual conversation in general can be conducted between strangers or friends. Furthermore, small talk topics are much broader than commonly assumed. Schneider [16] suggests three situation categories from which topics can be chosen during small talk, but points out that use of topics differs among, and even within, cultures. He proposes a sequence for discussing a topic during small talk, consisting of an initial question/answer pair, followed by several further turns of question/acknowledgment or idling behavior, often referred to as "Schneiders sequence".

The main purpose of casual conversations is the maintenance of social identities and relations. Eggins and Slade [6] state that, while there are no restrictions of topic selection in casual conversations, the structure of casual conversations is an important part of the process of creating and maintaining social roles. They consider *speech functions* to be a fundamental part of discourse structure and present a network of speech functions intended to be used to annotate and analyze casual conversations. While the authors have in mind conversations between friends or workmates when talking about casual conversations, the common understanding of small talk is one of discourse mainly occurring between strangers, or at least people that are not close friends.

In conclusion, even when small talk is about things which are not important, the skill of conducting small talk is important in that it helps to establish social relations. Thus small talk is more than "chit-chat" in the sense of idle talk.

2.2 Virtual Agent Research

The idea to use human-like dialogues was early adopted in recommender systems. Those systems where intended to operate in a closed task domain and therefore only task specific dialogue capabilities were implemented.

With the emergence of virtual agents the need for more elaborate dialogue systems arose. Bickmore and Cassell [4] introduce the idea of implementing small talk in their agent REA, an embodied conversational agent for a real estate sales domain. In addition to pursuing its task-oriented goal, REA tries to accomplish non-task-oriented, interpersonal goals by engaging the interlocutor in small talk. The interpersonal goals mainly serve the purpose to prepare the interlocutor for the next task-oriented dialogue move, by making him feel more comfortable and relaxed. In [5] Bickmore and Cassell identified user trust to be most important for their scenario. With the aid of small talk, the agent is enabled to affect this dimension by e.g. establishing common ground and conducting "facework". An activation network-based approach is used for discourse planning, allowing for a fine grained control of REA's conversational strategies. However, since user responses are mainly ignored, there seems to be no need for structuring the dialogues on utterance level.

Meguro et al. [13] use an HMM-based approach to compare so called listening-oriented dialogues to casual conversations. They successfully trained HMMs to distinguish between the two dialogue types. Analysis of the HMMs gave further insight into the structural differences of listening-oriented dialogues and casual conversations, namely the frequency of tags and the transitions between them. Novielli and Strapparava [14] use HMMs for automatic dialogue act classification of utterances. They exploit differences in dialogue patterns for categorizing different types of users.

Klüwer [10] criticizes that, despite the fact that small talk has been acknowledged an important part of human-agent conversations, no computational model has been developed, and even the most prominent annotation schemes for dialogue acts lack a specialized tag set for social acts. The author presents a set of dialogue acts intended as an extension for existing tag sets. Her analysis of a dialogue corpus reveals the occurrence of several different sequences during small talk, supporting her claim that the use of a single sequence (e.g. Schneider's sequence) may not be sufficient.

Endrass et al. [7] investigate cultural differences in small talk and evaluate their findings using virtual agents. Summarizing literature they state that small talk dialogues in Asian and Western culture differ in structure, in that Western small talk dialogues tend to be more sequential than the Asian ones. Using Schneider's sequence as a basis for their computational model of small talk, they plan to adapt the sequences according to the cultural background of the interlocutor.

In conclusion, while there are many indications in the current literature that more elaborate small talk capabilities for virtual agents would seem advantageous, research on this issue has only begun. We consider the structuring of dialogue in varied small talk sequences an important starting point for further progress.

3 Structuring Dialogue

To motivate our aim to provide a fine grained structure for small talk dialogues, consider the short, fictitious example dialogue depicted in Fig. 1. In fact, several different dialogues can be constructed by omitting certain utterances. E.g. sequences consisting of the utterances *1,3*; *1,2,3*; *1,3,4,5*; *1,3,6,7*; *1,2,3,4,5,6,7*, etc. all resemble short conversations that make perfect sense. However, the complexity of the sequences differ. One could even argue that they not only differ in terms of structure, but in the level of commitment they convey. In fact, if the whole conversation only consists of Question-Answer (QA) pairs (like the *1,3* sequence) the dialogue could be considered rather superficial.

```
1. A:  Do you like soccer?
2. A:  I mean do you like watching it on tv?
3. B:  Sometimes.

4. A:  Uhh hu.
5. B:  I sometimes do watch world cup matches. I am
       not that much into watching every game that's on.

6. A:  So, you don't like soccer?
7. B:  Well not that much!
```

Fig. 1. Short example of a dialogue

3.1 Present System Architecture

The architecture of our agent consists of a BDI interpreter, based on JAM [9]. Beliefs about the world are stored as facts in the agent's world knowledge. Actions of the agent are guided by his internal goals, intentions, and external events. The agent is able to sense his environment through cameras and microphones. Percepts received through these sensors lead to reactive and deliberative behaviors of the agent.

While the agent is able to generate synthetic speech, interlocutors use a keyboard for input in the present setting. Utterances received as input through the keyboard are processed in the agent's deliberative component. The conversational behavior of the agent is realized within the deliberative component, either as response to the utterance of an interlocutor or as proactive behavior.

In order to generate a response to an utterance, the keyboard input is processed in several steps within the deliberative component. The first step is the interpretation of the textual input. Pattern matching rules classify the input among general semantic concepts [11]. In a second step a communicative function of the utterance is determined, by again employing rules matching certain features of the input. The communicative function consists of three parts – the *performative*, *reference*, and *content* part – covering semantic and pragmatic aspects of the utterance [11].

The original input and the assigned communicative function are passed on to the dialogue engine. Within the dialogue engine an appropriate response is determined from a set of rules. To be more precise, the plan with the highest utility for the current goal, among all BDI plans constituting the agent's dialogue knowledge, is selected and executed. As a last step, the behavior planner generates a multi-modal utterance that is then performed by the agent.

Our agent is employed as a museum guide, thus main effort was put into the design of the agent's presentation capabilities. Therefore, task-oriented dialogue knowledge is structured into small units [11], while small talk capabilities of the agent are mainly restricted to simple keyword matching and direct responses, resulting in short QA sequences. However, both types of dialogue rely on the communicative function in order to determine subsequent utterances.

In the following we present our approach of extending the dialogue engine to allow for a fine grained control of small talk and small talk sequences. In our approach, structuring of dialogues takes place on two different levels within our dialogue engine. The micro level consists of meta information that dialogue acts can be tagged with. Ordering of different dialogue acts into sequences, and therefore dialogues, is conducted on the macro level.

3.2 Micro Level

The concept of tagging dialogue acts with meta information is already present in the dialogue engine in terms of the communicative function. Extending this, we introduce another meta information – the *speech function*. Figure 2 depicts an utterance of the example dialogue as represented in the dialogue engine. The two types of meta information are discussed in the following.

```
<act communicative_function="askFor.content.dislikesSoccer"
     speech_function="rejoinder.track.probe" >
  So, you don't like soccer?
</act>
```

Fig. 2. XML notation of a dialogue act of our agent annotated with meta information.

Communicative Function The *communicative function* meta information consists of three parts reflecting different information about the dialogue act of the agent and his interlocutors (see Fig. 3). Information about the speaker's intention is conveyed in the *performative* part. The *reference level* part determines the level of dialogue the act refers to: the *interaction*, *discourse*, and *content* level. The *content* part contains the semantic part of the dialogue act. E.g. it may contain the topic the utterance refers to (cf. [11]).

```
<performative>.<reference level>.<content>[arguments]
```

Fig. 3. Three independent parts constitute the communicative function of the dialogue engine [11].

Only two different types, *provide* and *askFor*, are distinguished in the performative part of the function. These two types correspond to the *giving* and *demanding* speech roles used by Halliday [8]. While being applicable for simple small talk consisting of QA pairs, the distinction of only two performatives in the communicative function is not suitable to structure the conversation on a fine-grained level.

Speech Function The *speech function* meta information can be considered an extension of the communicative function further specifying the action of the dialogue act. Halliday [8] suggests four basic speech functions, two for each speech role, to capture the commodity and role of dialogue initiating moves (cf. [6]), and eight corresponding responding speech functions. Eggins and Slade [6] provide a finer subclassification of Halliday's basic speech functions, in order to account for the more subtle structure of casual conversations. Their speech functions are classified among four subcategories as illustrated in Fig. 4.

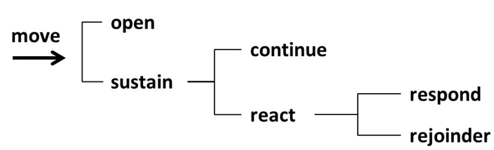

Fig. 4. Speech function network for casual conversations after [6]

As stated in Sect. 2.1 small talk is considered a subset of casual conversations. For this reason, we argue to exploit the speech functions for casual conversations in conversational agents that are to engage in more sophisticated small talk. In the initial implementation a subset of the speech functions suggested in [6] is used. An overview of the speech functions used in our dialogue engine is given in Table 1. In addition to the speech functions an example of an utterance, and its corresponding communicative function (if existent), is given. Note that some speech functions share the same communicative function.

3.3 Macro Level

The macro level is concerned with deciding how to select appropriate dialogue acts during conversation. One of the aims of using meta information is the

Table 1. Subset of speech functions taken from Eggins and Slade [6]

Move Type	Speech Function	Communicative Function	Utterance
open	attending	provide.interaction.greeting	Hey!
	offer	provide.discourse .offer.guessingGame	Shall we play a funny guessing game?
	statement	provide.content.weather	The weather is really nice today.
	question	askFor.content.likesSoccer	Do you like soccer?
continue	monitor	askFor.content.confirmation	You know?
	elaborate	askFor.content.likesSoccer	I mean do you like watching it on tv?
respond	register		Mmm
	support.reply	provide.content.confirmation	Right!
	confront.reply	provide.content.disagree	No, sorry!
rejoinder	check	askFor.content.who	Who?
	confirm	askFor.content.confirmation	Did he?
	probe	askFor.content.dislikesSoccer	So, you dont like soccer?
	counter	askFor.content.confirmation	Does this even matter?

reusability of generic utterances. E.g. utterances used as feedback channel, like "Uhh hu", can be used in a lot of situations, regardless of the content of prior utterances. On the other hand, related work (cf. Sect. 2.2) and the example given in Fig. 1 revealed that a variety of different sequences may occur within small talk conversations. Meguro et al. [13] demonstrated that even the type of conversation can be inferred from the transitions of dialogue acts.

As stated in Sect. 1, in some situations it may be crucial to reliably produce a certain sequence. E.g. in small talk training applications the agent should start with a very simple sequence, like Question-Answer. Over the course of training more complex sequences may have to be produced. Exploiting the introduced speech function meta information, a fine grained control of the course of conversations is possible. Figure 5 and Fig. 6 depict two possible sequences in our system, a simple QA, and a more complex sequence.

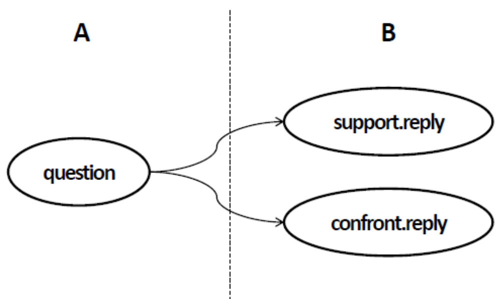

Fig. 5. Simple QA sequence. Dashed line represents the end of a turn.

The following two problems have to be solved in order to enable the dialogue engine to select an appropriate response to an interlocutor's utterance:

1. **Utterance tagging:** Assign an appropriate speech function to interlocutor's utterance
2. **Utility adjustment:** Determine probability values for speech function candidates and adjust utility values of corresponding plans

Utterance tagging When annotating dialogues between two participants, the speech function of an utterance can be determined by only referring to the previous speech function (cf. [6]). But, as stated above, the dialogue engine should be able to produce structured sequences like the one presented in Fig. 6. It is obvious, that in this case, it is not sufficient to rely on the immediate predecessor of the current utterance. In the first case, the *reply* utterance of B is followed by a *probe* utterance. In the second case, it is followed by a *reply* utterance.

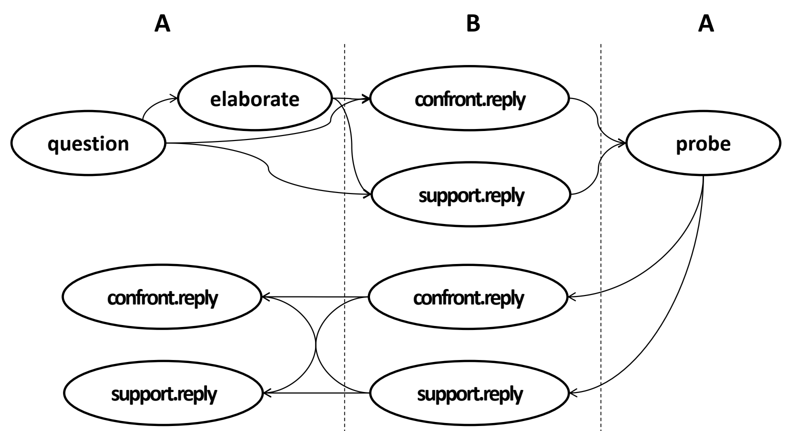

Fig. 6. Complex sequence. Dashed lines represent end of turns.

One option to assign speech functions to utterances would be to use a similar approach as used for deriving the communicative function, as described in [11]. However, this would require to provide dedicated rules for every possible sequence, resulting in a lot of redundant rules that only differ in the context of their sequence. Therefore, the process of assigning the speech function is done as a post processing step following the interpretation stage (cf. [11]) after every utterance of the agent's interlocutor. It is carried out by utilizing JAM's feature of employing Java methods. The Java part contains a representation of possible sequences, the sequence that is currently produced, the advancement in the current sequence, and a mapping of conversational to speech functions. The method for selecting the speech function is provided with the communicative function of the utterance that was determined during the interpretation stage. As output it returns the best matching speech function for the utterance. This way the communicative function is mapped to a corresponding speech function. By exploiting the sequence's history, utterances with similar communicative functions can be distinguished.

Algorithm 1 depicts the pseudo code for determining the speech function of an utterance. Given the communicative function `CF` and a set of speech functions `sequence_sfs` that are given due to the current position in the sequence, a set of possible speech functions is calculated. The speech function with the highest probability is selected and returned.

Algorithm 1 Pseudo code for selecting the speech function of an utterance

```
function SELECT_SF(CF : cfunction, sequence_sfs : {sfunction})
    possible_sfs = ∅

    for each sfunction SF with cfunction CF do
        possible_sfs = possible_sfs ∪ {SF}
    end for

    if possible_sfs ∩ sequence_sfs ≠ ∅ then
        return sfunction SF from (possible_sfs ∩ sequence_sfs)
                    with probability(SF) == max
    else
        return sfunction SF from possible_sfs
                    with probability(SF) == max
    end if
end function
```

Utility adjustment The BDI-based implementation of the dialogue engine allows for a flexible solution of providing alternatives for the agent's next dialogue contribution. Since utterances of our agent are represented within BDI plans, the utility values of these plans can be exploited to guide the agent's responses. The utility values define an order among plans. The plan with the highest utility value is selected as the most promising for fulfilling the current goal.

In contrast to the *utterance tagging* task described above, probabilities for possible following speech functions have to be determined after dialogue contributions of the agent and his interlocutor. Consider Fig. 6 with A being the agent. After A's first contribution A could try to hold the turn and continue with a further utterance. Accordingly, a probability value $\in [0, 1]$ is determined for every speech function known by the system (cf. Table 1) in the utility adjustment step after every utterance. To be accessible by the dialogue plans, the probabilities are stored as facts within the engine's dialogue knowledge.

Figure 7 depicts a simplified plan of our system. The utility value of the plan is multiplied by the corresponding speech function probability in the *precondition* part.

```
Plan:{
  NAME: "rule-0001 - continue.extend"
  GOAL: PERFORM match;
  PRECONDITION:
     (assign $util 10);
     (assign $util (* $util $continueextend));
     (FACT turn-holder "system");
  BODY:
    PERFORM collect-act
             (+ "<act sfunction=\"continue.extend\">
                   I mean do you like watching it on tv?
                 </act>");
  UTILITY: $util;
}
```

Fig. 7. The utility value of a BDI plan is altered according to the probability of its speech function.

4 Conclusion and Future Work

In this paper an approach of improving small talk capabilities of an embodied conversational agent was presented. Speech functions of human casual conversations are used to tag utterances on the micro level. On the macro level, the tagged utterances can be ordered into arbitrary sequences found in human-human dialogues. Using these speech functions, our extended dialogue engine is able to produce various dialogue sequences as introduced in the example dialogue in Sect. 3.

In order to determine if the use of speech functions to structure dialogues on a fine-grained level actually leads to enhanced interactions, an evaluation with human interlocutors is planned. One possibility considered is to integrate the enhanced dialogue engine in the museum setting the agent daily operates in. Pfeiffer et al. [15] provide information about the mean length of dialogues (in terms of time and utterances) between the agent and his visitors in the museum setting. Following [2], an improvement in dialogue length could be judged to indicate an overall improvement of the system's acceptance due to its increased small talk capabilities. An accompanying questionnaire will be used to obtain additional hints about the qualitative changes. Since only a subset of the speech functions suggested for casual conversations in [6] was used to demonstrate the possibility of integrating speech functions into an existing dialogue engine, results of the evaluation could be used to assess if the full set of speech functions for casual conversations is actually needed.

In our initial approach speech functions are determined relying on the communicative function. An improvement could be the use of a machine learning approach to assign speech functions as described in [14]. Another option would be to focus on a more linguistically motivated approach. Following Halliday, Eggins and Slade [6] make use of mood and modality of the grammatical realizations of moves to identify speech functions.

The importance of being able to produce structured sequences in certain applications was stressed throughout this paper. However, the possibility to adapt the choice of sequences is important for an agent that engages in elaborate small talk to improve the relationship with his interlocutors. Bickmore and Cassell [5] found that the acceptance of an agent using small talk may also depend on the interlocutor's personality. Eggins and Slade state that conversations with close friends tend to be more confronting than the ones we have with work colleagues, because "conversations between close friends involve as much probing of differences between friends as confirming the similarities which brought them together as friends in the first place" [6]. Consequently, in future work we will focus on how information supplied by a Person Memory (cf. [12]) of an agent can be further used to improve the small talk conversation in terms of dialogue structure.

5 Acknowledgments

This paper is a preprint version of an article published by Springer-Verlag. The original publication is available at http://link.springer.com/chapter/10.1007%2F978-3-642-33347-7_11.

References

1. Small talk (2012), http://dictionary.cambridge.org/dictionary/british/small-talk?q=small+talk, [Accessed Apr. 27, 2012]
2. Benyon, D., Hansen, P., Webb, N.: Evaluating human-computer conversation in companions. In: Proc. 4th International Workshop on Human-Computer Conversation (2008)
3. Benyon, D., Mival, O.: From human-computer interactions to human-companion relationships. In: Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia. pp. 1–9. ACM (2010)
4. Bickmore, T., Cassell, J.: Small talk and conversational storytelling in embodied conversational interface agents. AAAI fall symposium on narrative intelligence pp. 87–92 (1999)
5. Bickmore, T., Cassell, J.: Relational agents: a model and implementation of building user trust. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 396–403. ACM (2001)
6. Eggins, S., Slade, D.: Analysing Casual Conversation. Cassell (1997)
7. Endrass, B., Rehm, M., André, E.: Planning small talk behavior with cultural influences for multiagent systems. Computer Speech & Language 25(2), 158–174 (2011)
8. Halliday, M.: An Introduction To Functional Grammar. Edward Arnold, 2nd edn. (1994)
9. Huber, M.: JAM: A BDI-theoretic mobile agent architecture. In: Proceedings of the third annual conference on Autonomous Agents. pp. 236–243. ACM (1999)
10. Klüwer, T.: "I Like Your Shirt" - Dialogue Acts for Enabling Social Talk in Conversational Agents. In: Proceedings of the 10th international conference on Intelligent virtual agents. pp. 14–27. Springer (2011)
11. Kopp, S., Gesellensetter, L., Krämer, N., Wachsmuth, I.: A conversational agent as museum guide - design and evaluation of a real-world application. In: Intelligent Virtual Agents. pp. 329–343. Springer (2005)
12. Mattar, N., Wachsmuth, I.: Who Are You? On the Acquisition of Information about People for an Agent that Remembers. In: ICAART 2012 - Proceedings of the 4th International Conference on Agents and Artificial Intelligence. pp. 98–105. SciTePress (2012)
13. Meguro, T., Higashinaka, R., Dohsaka, K., Minami, Y., Isozaki, H.: Analysis of listening-oriented dialogue for building listening agents. In: Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue. pp. 124–127. No. September, Association for Computational Linguistics (2009)
14. Novielli, N., Strapparava, C.: Dialogue Act Classification Exploiting Lexical Semantics. In: Conversational Agents and Natural Language Interaction: Techniques and Effective Practices, pp. 80–106. Information Science Reference (2011)
15. Pfeiffer, T., Liguda, C., Wachsmuth, I.: Living with a Virtual Agent: Seven Years with an Embodied Conversational Agent at the Heinz Nixdorf MuseumsForum. Group (2011)
16. Schneider, K.P.: Small talk: analysing phatic discourse. Hitzeroth (1988)