Towards Computational Models for a Long-term Interaction with an Artificial Conversational Companion

Sviatlana Danilava[1], Stephan Busemann[2], Christoph Schommer[1] and Gudrun Ziegler[1]

[1]*University of Luxembourg, 6 Rue Coudenhove-Kalergi, Luxembourg, Luxembourg*

[2]*Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH, Stuhlsatzenhausweg 3, Saarbrücken, Germany*

sviatlana.danilava@uni.lu, stephan.busemann@dfki.de, christoph.schommer@uni.lu, zieglergudrun@gmail.com

Keywords: Artificial Companions, Models of Interaction, Human-Computer-Interaction, Instant Messenger

Abstract: In this paper we describe a design approach for an Artificial Conversational Companion according to earlier identified requirements of utility, adaptivity, conversational capabilities and long-term interaction. The Companion is aimed to help advanced learners of a foreign language to practice conversation via instant messenger dialogues. In order to model a meaningful long-term interaction with an Artificial Conversational Companion for this application case, it is necessary to understand how natural long-term interaction via chat between human language experts and language learners works. For this purpose, we created a corpus from instant messenger-based interactions between native speakers of German and advanced learners of German as a foreign language. We used methods from conversation analysis to identify rules of interaction. Examples from our data set are used to illustrate how particular requirements for the agent can be fulfilled. Finally, we outline how the identified patterns of interaction can be used for the design of an Artificial Conversational Companion.

1 INTRODUCTION

The vision of artificial agents that *"listen to spoken sentences"* and are *"nice personalities"* and our *"little electronic friends"* is not new, see for example (Winograd and Flores, 1987). In 2005, Y. Wilks introduced the notion of an *Artificial Companion* as *"an intelligent and helpful cognitive agent which appears to know its owner [...], chats to them [...], assists them with simple tasks"* (Wilks, 2005). The agent must be able to maintain a sustained discourse over a long time period, serve interests of the user and have a lot of personal knowledge about the user (Wilks, 2010).

The idea to use conversational agents (chatbots) in second language ($\mathscr{L}_2$) acquisition is long of interest (Kerly et al., 2007; Zakos and Capper, 2008). Language acquisition *"requires meaningful interaction in the target language [...] in which speakers are concerned not with the form of their utterances but with the messages they are conveying and understanding"* (Krashen, 1981). Predictability of responses, lack of personality and inability to remember the interaction history are reported as shortcomings of current agents for conversation training (Shawar and Atwell, 2007). Agent designers battle against these issues by creating more sophisticated patterns for domain-unrestricted language understanding, storing information about the user and already used responses (Jia, 2009). However, the design of the agents is still focused on the content of responses, but not on language as a co-constructed meaningful action.

We consider the application scenario where advanced learners of a foreign language practice conversation in dialogues with an Artificial Conversational Companion (ACC). In earlier work, we identified the minimum requirements that an artificial agent must satisfy in order to be mentioned as an ACC (Danilava et al., 2012). We refined the requirements for the application scenario of conversation training in $\mathscr{L}_2$-acquisition. We focused on interaction via instant messenger (IM) because it combines the advantages of spoken and written communication being conceptually oral and medially written (Koch and Oesterreicher, 1985).

In this paper, we describe our ACC design approach based on empirical data from IM dialogues according to the earlier identified requirements of conversation (natural language understanding and generation (NLU/NLG), cognitive abilities, emotional competence, socio-cultural competence), utility, adaptivity and long-term interaction. We see a long-term interaction as a process of construction of longer dialogues over a stretched period of time, where the length of each dialogue and the number of dialogues do not have any predefined minimum value. It remains open, when the interaction ends.

We extract patterns from dialogues between humans that will help make an interaction with an ACC close to a natural interaction that is co-constructed by all the participants as a meaningful activity according to rules of social interaction and by means of the selected communication medium.

We explain the methodology and briefly describe the data set in Section 2. We illustrate our design approach on examples from the dataset in Section 3 followed by conclusions in Section 4.

2 METHOD

In order to model a long-term interaction with an ACC via IM dialogue, it is necessary to understand how natural long-term IM-based interaction between human language experts and language learners works. We created and used data from natural interactions for this type of analysis. Language experts provided interaction patterns for the future ACC, and language learners offered information for user modelling.

2.1 Data

IM chat is subject to intensive research in conversation analysis (Orthmann, 2004; Nardi et al., 2000), computer-mediated collaborative work (Jiang and Singley, 2009) and natural language processing (Forsythand and Martell, 2007). The data sets used come from natural workspace interaction (Avrahami and Hudson, 2006) or interaction experiments (Solomon et al., 2010). These data sets are not available as a resource for the research community. Corpora from multi-user open chatrooms are available for research, see for example (Lüdeling, 2009; Lin, 2012). They do, however, not satisfy the requirements of dialogic interaction between language experts and language learners over a long period of time. For this reason, we created our own corpus of IM based expert-learner dialogues for German as a focus language. Space limitations prevent us to offer more than a concise description of the data set.[1]

Voluntary participants – 4 German native speakers and 9 advanced learners of German as $\mathscr{L}_2$ – communicated within 4-8 weeks via IM. Each chat session took between 20 and 90 minutes. The parties communicated with the same partner for the complete duration of the experiment. The participants produced a total of 72 dialogues, which correspond to ca. 2.500 minutes of IM interaction, ca. 4.800 messages of 10 tokens average length, ca. 52.000 tokens in total, and ca. 6.100 unique tokens.

Since the data set is quite small, quantitative methods are not applicable to achieve statistically significant results. However, qualitative methods from ethnomethodology and conversation analysis can deliver reliable results in understanding rules of interaction based on small-scale data sets. Examples from the data set taken from dialogues with different participants pairs illustrate the most important types of rules. We use the notation N for experts and L for learners.

2.2 Modeling Approach

Top-down ACC design approaches according to abstract requirements suffer in general from the impossibility to foresee all potential ways for the interaction flow. In contrast, a bottom-up data driven design could lead to a huge number of abstract classes with unclear relations between them and make the system design unmanageable. We combine the top-down requirements for ACCs with the bottom-up approach commonly used in conversational analysis.

Conversational analysis usually is performed in three steps (cf. e.g. (Orthmann, 2004)):

1. Looking through the data without preconceptions about what may be found.
2. Pattern synthesis: more abstract, generalised description of structures found.
3. General abstract description of interaction rules.

In particular, we focused on rules in interaction, where the participants make the meaningful activity and the social interaction (social closeness or distance) explicit. There are dialogue patterns specific for learners (e.g. different types of errors) and experts (e.g. error correction), and in patterns that disclose disruptive factors in interaction as, for example, too long response time. Finally, we outline how the detected patterns serve as a basis for computational models of long-term interaction in general and for design of an ACC for conversation training in particular.

3 DATA ANALYSIS

3.1 Patterns for Utility

As reported by the volunteers, the motivation to participate in the experiment was their willingness to

[1] Visit *http://wiki.uni.lu/mine/Sviatlana+Danilava.html* for a detailed description of the data collection.

help the organiser, willingness to improve their language skills and conversational competence, practice conversation in German, willingness to do something new, to get in contact with people from other countries and not to be passive. Furthermore, we observed in the data two actions focused on language learning: error correction and explanation of new lexical material. These activities provide valuable patterns for both, utility and NLU/NLG.

3.1.1 Awareness of the Meaningfulness

Helping the organiser is documented by participants especially in closing sequences, where they "did their job". Example 1 shows participants' awareness in a particular interaction sequence: N says "ok, we have produced a lot of text" expressing his personal interpretation of the meaning of this interaction. Similar sequences also occur in dialogues produced by other pairs.

Example 1: Talking about the meaningful activity.

Time	Sndr	Message Body
18:44:44	N	ok, wir haben jetzt eine große Menge Text produziert. Es war sehr schön mit Dir zu plaudern. Fällt Dir noch was ein?

Example 2 illustrates how participants demonstrate that they "did their job" for the whole experiment (We explain the use of parentheses in Sec.3.4).

Example 2: Talking about activity completion.

Time	Sndr	Message Body
21:56:18	N	Ich habe übrigens vorgestern mit unserer Organisatorin gesprochen. Wir haben unserer Gesprächsanzahl heute erfüllt :-)
21:57:28	L	Ich hoffe, dass unser Chat für sie nutzbar ist)))

After each pair of participants completed 8 dialogue sessions, they could choose to keep communicating or to abandon the chat. A task like conversational training can also be finished after a certain number of sessions. A possible scenario for an ACC is to inquire with the user from time to time if the interaction should continue or come to an end.

3.1.2 Error Recognition and Error Correction

Learners often produce ungrammatical sentences. For an ACC, even pattern-based NLU is challenging due to incorrect use of lexical items. Therefore, error models are important for utility and are an essential part of NLU. There are error-tagged corpora for learners' language. However, they are created from conceptually written data (usually essays) and do not contain corrections (Lüdeling et al., 2005; Boyd, 2010).

Typical errors depend among others on the level of language proficiency, and on ones's native language. Statistical error models offer a basis for grammar and style error recognition. However, they are usually based on native speakers' data and cannot deal with wrong use of lexical items (Crysmann et al., 2008). Automatic error recognition is problematic because of contextual factors rendering an otherwise grammatical expression invalid. In Example 3, N corrects "bin ich frei" even though this is a grammatical sentence.

Example 3: Error correction.

Time	Sndr	Message Body
19:41:15	L	Zur Zeit bin ich frei um Diplomarbeit zu schreiben
19:41:58	N	Falls ich das korrigieren darf: Du "hast" frei, nicht "bist". :-)
19:42:48	L	Danke
19:43:16	L	Ich habe frei :)

Besides explicit embedded repair sequences as in Example 3, the data set also exhibits error corrections, for example, in the form of indirect repairs (Example 7: "2 Teste" - "2 Tests") or a direct correction response to a erroneous sentence. The choice of the correction form depends on the level of social interaction and the error type. There is a conflict between the learners' expectation that the native speakers help them to improve their skills by correcting errors and the native speakers' wish not to be boring and therefore not to correct too much.

3.1.3 Models for Conversational Training Task

German participants introduced idiomatic expressions, which they thought the partners did not know, and explained the meaning (Example 4).

Example 4: Introducing new lexical material.

Time	Sndr	Message Body
20:01:24	N	Was macht die Kunst?
20:03:29	L	Was bedeutet dieser Ausdrück? Ich verstehe nicht (((
20:04:09	L	Meinst du Tanzen?
20:04:44	N	Das habe ich mir schon gedacht :-) `[explaining]`

German participants explained the meaning of some words in form of jokes (Example 5).

Example 5: Introducing and explaining multiple meanings of *treffen*.

Time	Sndr	Message Body
20:32:40	N	Ich finde Witze mit doppelter Wortbedeutung ganz lustig.
20:33:02	N	Ich probiers mal, vielleicht verstehst du ihn.
20:33:32	N	Im Wald treffen sich zwei Jäger, beide tot.
20:35:33	L	Ich brauch deine Hilfe. Gib mir bitte einen Fingerzeig!
20:37:33	N	Okay :-) `[explaining]`

Lexical error correction or explanation of unknown words can be implemented with external resources for idioms and meaning explanations, including but not limited to Wikipedia or online dictionaries.

3.2 Patterns for NLU/NLG

Conversation is co-constructed by all participants. It contains at least a content part (e.g. topic) and a management part where the participants display that they are talking and intend to keep doing so. Responsiveness is an observable phenomenon for the interaction management. A huge amount of research is focused on the the content part. To our knowledge, responsiveness has never been taken into account for chatbot design. IM responsiveness is influenced by uncontrolled elements like parallel activities of the participants, typing speed, network delays, experience in chat interaction. Time stamp and the length of the received message and the produced response are, however, analysable for both, researchers and participants.

Interaction parties share the understanding of what is acceptable. In Example 6, L replied to N's question. More than 5 minutes later L posts a request whether N still has time for chat, displaying that the time interval is too long. N replies with an excuse offer, displaying the awareness of the acceptable time interval.

Example 6: The Length of the allowed time interval

Time	Sndr	Message Body
17:46:42	N	wie bist Du zu diesem Chat-Projekt gekommen?
17:47:14	L	meine Lektorerin hat mir gesagt. [explaining]
17:52:33	L	[N_firstname], wenn du schon keine Zeit hast, dann schreibe dann bitte) sonst kann ich sehr lange schreiben)))
17:53:08	N	oh, Verzeihung! bin wieder da, sorry!
17:54:40	L	ok)

Models for presence requests and back-channeling can be described according to responsiveness criteria. The initial value of acceptable responsiveness can be configured according to average values from interactions between native speakers, and then adapted according to the learner's behavior. Responsiveness-based rules can be defined for repairs, self-repairs, repetitions and interaction management at unit (topic and action) boundaries.

3.3 Required Cognitive Abilities

The participants did not know anything about their partners prior to the experiment. Initially, they asked their partners about their names, age, locations and occupation. This knowledge might be required for initialising appropriate social closeness/distance in interaction (see also Sec. 3.5). Some facts remain important for a longer period of time (e.g. names, or locations), some can be put aside after a particular time (e.g. examinations), or maybe forgotten.

Once some facts are learned, it is appropriate to ask about the state of the fact later. For example, when some of the participants told that they have exams, their partners asked about the results later (Example 7), which also would be an appropriate reaction for the ACC.

Example 7: Simple inferences: talking about significant events (23. May and 31. May)

Time	Sndr	Message Body
18:29:10	L	leider auch nicht(((morgen schreibe ich 2 Teste in Deutsch und Englisch. und wie du verstehst, habe ich noch nicht sie gelernt=))
18:30:02	N	2 Tests? ok, klar, dann erstmal viel Erfolg dabei!
18:29:23	N	hey, wie waren Deine Prüfungen?
18:35:21	L	[...] alle Tests wurden schon SEHR gut geschrieben. [...]

Similar patterns exist for football games (the experiment took place during a European soccer championship) and can be generalised for all significant events. This can be implemented as an inference rule that takes an important event, its date and information about expected results into account. The ACC may ask about results later.

3.4 Patterns for Emotional Competence

Human participants use paralinguistic and prosodic cues to show emotions in IM messages. They are expressed through capitalisation, spelling, punctuation or timing (Orthmann, 2004). The most analysed resource for displaying emotions in chat are emoticons (e.g. ":-)") and emotive language (e.g. "haha", "hihihi"). There are also reactive tokens (e.g. news markers "oh, wirklich?", "oh, nein!") expressing emotions.

Emotional competence of the agent concerns emotion recognition, interpretation and response generation. There are the same problems in correct emotion recognition and interpretation as for lexical items. In a multi-cultural dialogue, culture-specific items are used in addition. In Example 7, "(((" and "))" represent "sadness" and "joy", respectively. These symbols are ambiguous to opening and closing parentheses. This use of parentheses to display joy or sadness is typical for the learners, but was never used by the experts.

In Example 6, L makes a presence request containing ")" and ")))", which we interpret as an indicator of politeness. After the response of her partner N, L accepts his excuse by "ok)". In both cases, smiles cannot be interpreted as an indicator of joy.

Patterns for the use of reactive tokens, emoticons and emotive language in different actions – for example, corrections, presence requests, making appointments, initiating ending sequences, and marking unit boundaries – can be extracted from the data set.

3.5 Understanding Social Interaction

The participants use the form of address, the greeting forms, emotions, politeness, variations in the lexicon size or syntax to display social and emotional closeness or distance (Koch, 1994). The meaning of these resources can be interpreted differently by representatives of different cultures. However, there must be a set of universal rules for social interaction, as otherwise people would not be able to manage any intercultural communication without prior training.

We illustrate how the participants use the form of address: "Sie" (formal, third person plural) vs. "du" (familiar, second person singular) to find out their degree of social closeness in dialogues. We found patterns for explicit or implicit negotiation, whereby the implicit negotiation can take place within one interaction or stretch over multiple dialogues.

Example 8 shows an explicit negotiation: L asks, if it is good to write using "du", N answers that "it would be strange to chat using "Sie"".

Example 8: Explicit Negotiating Social Closeness.

Time	Sndr	Message Body
17:18:08	L	Und ist es gut, wenn ich "auf Du" schreibe?
17:19:13	N	oh, hab gar nicht gefragt... natürlich, auf "Sie" chatten wäre irgendwie seltsam :-)
17:19:48	L	ok=)

Example 9 describes an implicit negotiation: L starts with "Sie" and changes the form of address to "du" in her second turn. Her second turn repeats the question from her first turn ("I don't know your [=Sie] name.") but reformulates it using "du" ("What's your [=du] name?").

Example 9: Implicit Negotiating Social Closeness.

Time	Sndr	Message Body
19:57:31	L	Hallo! Entschuldigung, Ich weiß nicht, wie heißen Sie. [...] **mit freundlichem Gruß, `L_firstname`, `L_lastname`,** [...]!
19:59:57	N	Hallo `L_firstname`, das ist überhaupt kein Problem! Ich hoffe, alle Probleme sind gelöst und wir können ein bisschen chatten.
20:01:58	L	Ja, natürlich! wie heißt du?

To determine the appropriate degree of social closeness, the participants searched for similarities between them and their partners (age, location, occupation). At the current stage of research, it is not clear, which level of social interaction is appropriate for human-machine dialogues. It cannot be taken for granted that learners would use "Sie" in the interaction with a machine, but the machine could start with the polite form. In addition to the forms of address, use of reactive tokens, humour and sharing private information are examples of interaction phenomena, where social interaction is analysable.

3.6 Adaptivity Mechanisms

Similar to the book suggestion system described in (Rich, 1979), the ACC must be able to build individual user models from a very low amount of personal knowledge in the first conversation. As we illustrated in Section 3.3, information about age, gender, location and occupation of the participants was sufficient to initialise the level of social interaction for humans. The amount of knowledge about the partners increased over time. This knowledge can be used by the system in highly adaptive user models.

Adaptivity and anticipation may affect the single sessions (adaptivity within one interaction, for example topic sequences, responsiveness and user's mood) and the complete interaction history (for example preferred topics, changes in user's lexicon, error tracking, topic selection according to user's interests). The adaptation of a Companion's language (lexicon, style) to the user's language is not a general goal, because we cannot take it for granted that the learners use the foreign language correctly.

There are mutual dependencies between emotional, social and cultural aspects of the ACC design, which need to be explicitly modelled to trigger the adaptivity mechanisms correctly. For example, resources used for social interaction need to be selected according to social closeness. Emotions are embedded into the context of activity and are also adapted to the social interaction according to social closeness or distance. Telling jokes (Example 4), using colloquial expressions or showing emotions when something is not clear (Example 3) may be misplaced in an interaction with a large social distance.

3.7 Chances for Long-term Interaction

A long-term interaction with an ACC cannot be enforced. Potential users of an ACC – the learners – reported that they are curious to chat with the system for the first time, but they will not keep interacting with the system if it does not make sense for them. The goal of the ACC designers is to create necessary conditions to make long-term interaction possible. However, it cannot be a system requirement to achieve a dialogue of a particular duration in multiple sessions.

As Examples 1 and 2 show, the interaction parties are aware of them engaging in a joint activity meaningful for both of them, that this activity will span multiple sessions, and that after a particular number of sessions of hopefully pleasant interaction, the activity is completed, or its meaning has changed and they keep interacting. This was not avoidable for the data collection with volunteers. The interaction with an ACC does not necessarily have to end after a particular number of conversations.

We cannot take it for granted that the learners will accept the ACC as a language expert. The users will try to find out what the ACC does not understand. The ACC designers need to anticipate as much as possible of such scenarios in order to let the machine look smart or funny, but still polite according to the level of social closeness.

4 CONCLUSIONS

We collected data from human-human IM dialogues that revealed valuable patterns for social interaction and activity in the context of conversational training. In contrast to most existing Companion prototypes, we use empirical data for the design of an ACC. We described how we use these empirical data in order to satisfy the requirements of conversation, utility, adaptivity and long-term interaction.

To effectively use the patterns for the ACC design, a consistent, holistic rule framework will account for the interdependencies in recognising patterns in real-time during the interaction and in producing appropriate system (re-)actions in terms of utterance content and interaction management.

ACKNOWLEDGEMENTS

We would like to thank all the participants of the data collection experiment for their voluntary work.

REFERENCES

Avrahami, D. and Hudson, S. (2006). Responsiveness in instant messaging: Predictive models supporting interpersonal communication. In *HCI*, pages 731–740.

Boyd, A. (2010). EAGLE: an error-annotated corpus of beginning learner german. In *Proc. of LREC*. ELRA.

Crysmann, B., Bertomeu, N., Adolphs, P., Flickinger, D., and Klüwer, T. (2008). Hybrid processing for grammar and style checking. In *Proc. of the 22nd Int. Conf. on Comp. Linguistics*, pages 153–160. Coling.

Danilava, S., Busemann, S., and Schommer, C. (2012). Artificial Conversational Companion: a requirement analysis. In *Proc. of ICAART'12*, pages 282–289.

Forsythand, E. N. and Martell, C. H. (2007). Lexical and discourse analysis of online chat dialog. In *Proc. of the International Conference on Semantic Computing*, ICSC '07, pages 19–26. IEEE Computer Society.

Jia, J. (2009). CSIEC: A computer assisted english learning chatbot based on textual knowledge and reasoning. *Know.-Based Syst.*, 22(4):249–255.

Jiang, H. and Singley, K. (2009). Exploring bilingual, task-oriented, document-centric chat. In *Proceedings of GROUP '09*, pages 229–232.

Kerly, A., Hall, P., and Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Know.-Based Syst.*, 20(2):177–185.

Koch, P. (1994). Schriftlichkeit und sprache. In *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung*, pages 587–604. Walter de Gruyter.

Koch, P. and Oesterreicher, W. (1985). Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In *Romanistisches Jahrbuch*, volume 36. Walter de Gruyter.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Lin, J. (2012). *Automatic Author Profiling of Online Chat Logs*. Kindle Edition.

Lüdeling, A. (2009). *Corpus Linguistics: An International Handbook*. Mouton de Gruyter.

Lüdeling, A., Walter, M., Kroymann, E., and Adolphs, P. (2005). Multi-level error annotation in learner corpora. In *Corpus Linguistics*.

Nardi, B. A., Whittaker, S., and Bradner, E. (2000). Interaction and outeraction: instant messaging in action. In *Proc. of the ACM conf. on Computer supported cooperative work*, pages 79–88.

Orthmann, C. (2004). *Strukturen der Chat-Kommunikation: konversationsanalytische Untersuchung eines Kinder- und Jugendchats.* PhD thesis, Freie Universität Berlin.

Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3:329–354.

Shawar, B. A. and Atwell, E. (2007). Chatbots: Are they really useful? *LDV-Forum*, 22(1):29 – 49.

Solomon, J., Newman, M., and Teasley, S. (2010). Speaking through text: the influence of real-time text on discourse and usability in im. In *Proc. of the 16th ACM int.l conf. on Supporting group work*, pages 197–200.

Wilks, Y. (2005). Artificial companions. *Interdisciplinary Science Reviews*, 30(2):145 – 152.

Wilks, Y. (2010). Is a companion a distinctive kind of relationship with a machine? In *Proc. of the 2010 Workshop on CDS*, pages 13–18. ACL.

Winograd, T. and Flores, F. (1987). *Understanding Computers and Cognition: A New Foundation for Design*. Addison-Wesley Longman Publishing Co., Inc.

Zakos, J. and Capper, L. (2008). CLIVE - an artificially intelligent chat robot for conversational language practice. In *Proceedings of the 5th Hellenic conference on Artificial Intelligence: Theories, Models and Applications*, pages 437 – 442. Springer-Verlag.