Why are you Silent? - Towards Responsiveness in Chatbots

Sviatlana Danilava
University of Luxembourg
6 Rue Coudenhove-Kalergi,
L-1359, Luxembourg
sviatlana.danilava@uni.lu

Stephan Busemann
DFKI
Saarbruecken, Germany
stephan.busemann@dfki.de

Christoph Schommer
University of Luxembourg
6 Rue Coudenhove-Kalergi,
L-1359, Luxembourg
christoph.schommer@uni.lu

Gudrun Ziegler
University of Luxembourg
6 Rue Coudenhove-Kalergi,
L-1359, Luxembourg
zieglergudrun@gmail.com

CHI'13, April 27 – May 2, 2013, Paris, France.
Avec le Temps! Time, Tempo, and Turns in Human-Computer Interaction. Workshop at CHI 2013.

Abstract
In this position paper we consider temporal phenomena in interaction with text-based conversational agents. In particular, we focus on two dimensions of time in instant messaging dialogues: responsiveness as a measure for interaction placed in time, and interaction management performed by interaction participants caused by partner's exceeding of the maximum expected responsiveness.

Author Keywords
Responsiveness, Long-term Human-computer Interaction, Conversational Agents

ACM Classification Keywords
I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence - Intelligent agents; I.6.7 [Simulation and Modelling]: Model Development

Introduction
Chatbots fascinate researchers and practitioners from over the world [5]. Although responsiveness is a topic of intensive research in Computer-mediated Communication [1] and Spoken Dialogue Systems [6], it did not get so much attention in the area of chatbot design. Chatbots have been studied in the context of conversation training in second language acquisition (SLA). AIML-based bots have been extended by grammar rules to overcome the

Definitions

Adjacency pair: basic unit of interaction sequence organisation, composed at least of two turns produced by different speakers and relatively ordered in:

- **FPP** First pair part, initiates some exchange or action,
- **SPP** Second pair part, is responsive to an action [7].

Increment: a turn containing additional information to the previous turn produced by the same participant (Fig 3 turns 38 and 39 are increments of 37).

limitations of pattern-based language understanding [3]. Chat-based communication between the learners and the agent in SLA context has the advantages of spoken communication being conceptually oral, but it helps to overcome problems with learner speech recognition being medially written [4]. Since SLA usually takes a lot of time, the assisting agent needs to provide the necessary capabilities for a log-term interaction [2]. In particular, the agent needs to handle responsiveness issues and deal with increments produced by the user. We would like to address the following questions in this position paper:

1. Time intervals between parts of an adjacency pair that can be interpreted as non-talking under certain interactional conditions.
2. The role of time for parallel productions of turns.
3. Responsiveness-based recognition of increments.

We analyse a data set from human-human instant messaging (IM) dialogues to develop computational models of interaction that take into account responsiveness values among other factors.

413	L: Guten morgen, N_FirstName. bist du da?? =))	07:15:20
414	N: jetzt ja!	07:47:49

Figure 1: Are you there? Establishing the presence awareness: waiting longer than 30 minutes.

182	L: wie hast du diese Tage verbracht?	[Start action]	20:14:38
183	L: Warum schweigst du?	[Presence request]	20:17:01
184	N: Ach bei uns regnet es seit den letzten drei Tagen häufig. Deshalb konnte ich draußen nicht soviel machen. Morgen besuche ich einige Freunde. Wir werden am Samstag gemeinsam das erste Dtld-Spiel schauen.		20:17:25
185	N: Weil ich geschrieben habe ;-)		20:17:31

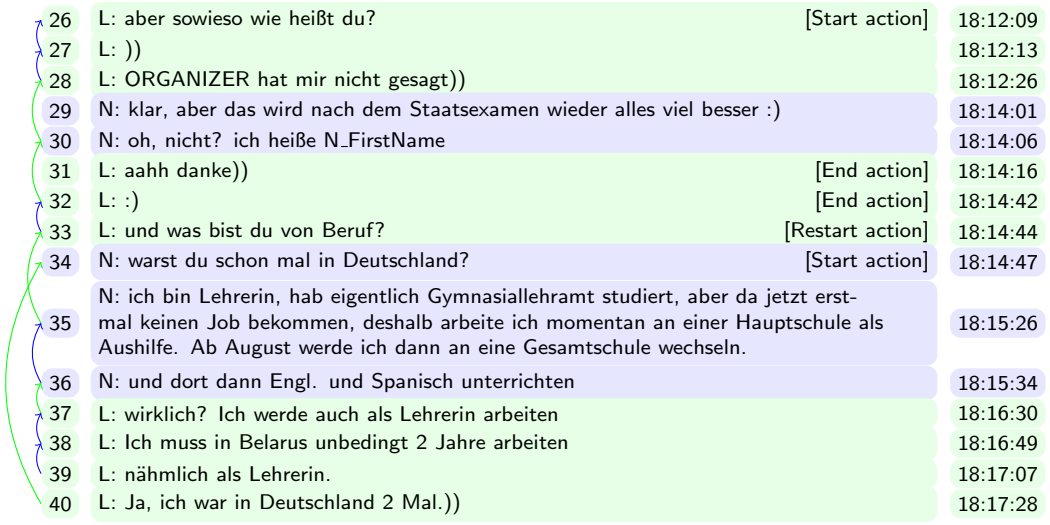
Figure 2: Why are you silent? In turn 183, L posts the request after 143 seconds of silence.

Responsiveness and turn-taking in IM

We created a corpus of IM dialogues between 9 advanced learners of German and 4 German native speakers. Volunteers produced 72 dialogues (ca. 4.800 messages, 6.100 unique tokens and 52.000 tokens in total). The parties communicated with the same partner for 4-8 weeks in chat sessions of 20-90 minutes[1]. Each message contains a server time stamp. The participants did not see each other directly, the communication was established over a forwarding chatbot which was always "available". Typing notifications and status changes were not visible for the parties. Thus, the awareness of co-constructing an interaction as a joint activity was only possible through posting messages. Figures 1, 2, 3 and 4 contain examples from four different IM sessions with different pairs of participants (L - learners, N - native speakers). Each turn contains a turn number (left), speaker (L or N), message body with some action mark-up (in square brackets, see page 4 for explanation) and time stamp (right). We consider each posted message as one turn. Each IM interaction in our data set is a sequence of messages produced by one pair of participants and ordered by time stamp. This order does not correspond to the interactional order among the messages: two neighbouring turns do not necessary form an adjacency pair. Edges between turns specify interactional connection between turns: green if both turns are produced by the same speaker (self), blue if turns are produced by different speakers (*other*).

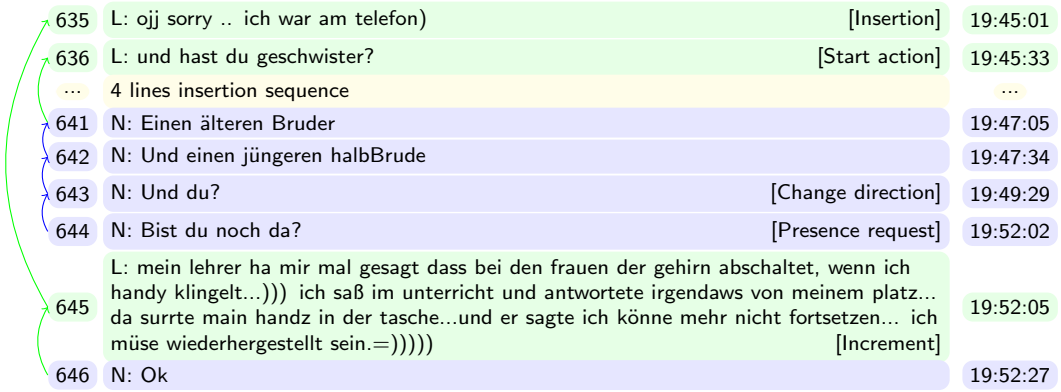
IM interaction (human-human and human-machine) is organised sequentially by means of turn-taking similar to spoken interaction as analysed in [7]. The possibilities for turn-taking in IM are determined by the tool and common turn-taking system inherent for human communication.

[1] The reader can find a detailed description of the data collection at *http://wiki.uni.lu/mine/Sviatlana+Danilava.html*.



26	L: aber sowieso wie heißt du?	[Start action]	18:12:09
27	L:))		18:12:13
28	L: ORGANIZER hat mir nicht gesagt))		18:12:26
29	N: klar, aber das wird nach dem Staatsexamen wieder alles viel besser :)		18:14:01
30	N: oh, nicht? ich heiße N_FirstName		18:14:06
31	L: aahh danke))	[End action]	18:14:16
32	L: :)	[End action]	18:14:42
33	L: und was bist du von Beruf?	[Restart action]	18:14:44
34	N: warst du schon mal in Deutschland?	[Start action]	18:14:47
35	N: ich bin Lehrerin, hab eigentlich Gymnasiallehramt studiert, aber da jetzt erst-mal keinen Job bekommen, deshalb arbeite ich momentan an einer Hauptschule als Aushilfe. Ab August werde ich dann an eine Gesamtschule wechseln.		18:15:26
36	N: und dort dann Engl. und Spanisch unterrichten		18:15:34
37	L: wirklich? Ich werde auch als Lehrerin arbeiten		18:16:30
38	L: Ich muss in Belarus unbedingt 2 Jahre arbeiten		18:16:49
39	L: nähmlich als Lehrerin.		18:17:07
40	L: Ja, ich war in Deutschland 2 Mal.))		18:17:28

Figure 3: Responsiveness and sequential organisation: the response in line 40 comes 161 seconds after the question.



635	L: ojj sorry .. ich war am telefon)	[Insertion]	19:45:01
636	L: und hast du geschwister?	[Start action]	19:45:33
...	4 lines insertion sequence		...
641	N: Einen älteren Bruder		19:47:05
642	N: Und einen jüngeren halbBrude		19:47:34
643	N: Und du?	[Change direction]	19:49:29
644	N: Bist du noch da?	[Presence request]	19:52:02
645	L: mein lehrer ha mir mal gesagt dass bei den frauen der gehirn abschaltet, wenn ich handy klingelt...))) ich saß im unterricht und antwortete irgendaws von meinem platz... da surrte main handz in der tasche...und er sagte ich könne mehr nicht fortsetzen... ich müse wiederhergestellt sein.=)))))	[Increment]	19:52:05
646	N: Ok		19:52:27

Figure 4: Are you still there? In turn 644, N posts the request after 153 seconds of silence.

The interaction is influenced by the tool (e.g. overlaps in production and parallel activities while chatting) and by human factors (e.g. general experience in IM, typing pace and language proficiency). Some of these factors are directly observable (e.g. time stamp and message length), some of them are inferable from the time stamp or message content (e.g. parallel production and telling about getting distracted by the girlfriend or a soccer match). In addition there are non-observable uncontrolled factors influencing the responsiveness, for example network delays, using different types of devices, parallel activities and learners' language proficiency. However, "the organization of turn-taking provides a way [...] to say non-trivially that someone in particular is not speaking, when in fact *no one at all* is speaking"[7, p.19]. We can determine "*who is not talking* and *what kind of talk they are not doing*" [7, p.20]. The former is possible due to the turn-taking system, the later is determined by adjacency pair organisation (we see, what kind of SPP is missing). The recognition of non-talking according to these two conditions is performed by the participants based on time of silence under particular interactional circumstances. Compare for example Fig. 2 turns 182, 183, and Fig. 1: more then 30 minutes silence in the later example do not lead to an interactional conflict, but 143 seconds of silence in the former example do (also in Fig. 4 turn 643, 644). However, long time intervals between turns 636 and 641 (Fig. 4) and turns 34 and 40 (Fig. 3) are not problematic, because the interaction does not "freeze" in that time. Turns 641 and 40 are referred to as late response, it is still a challenging task to recognise interactional connections for such turns automatically.

An example of parallel production is shown in Fig 3 turns 33 and 34: both parties select themselves as speaker. Why is this possible and acceptable in IM interaction? It

Mark-up

Start action the turn is an FPP and the first turn of a new activity.

Restart action the turn is an FPP but it belongs to an already introduced activity, possibly closed.

End action the turn is a possible end of an activity.

Change direction the turn holds all the interaction constraints introduced by the previous speaker but changes the addressee.

Presence request interaction management activity concerned with non-talking.

Insertion side turn sequences embedded into the main action of an interaction.

is allowed by the tool, determined by the turn-taking system, but also it is time to speak.

Responsiveness-based interaction profile

How the parties know, that it is time to speak? At the beginning of the data collection, the participants did not know anything about their partners, especially, about partner's typing pace, internet connection speed, device used and language proficiency. All of these factors influence message production speed and thus responsiveness. To obtain this information from the interaction, we look at the time intervals between increments. The first insights into data show, that the time intervals in *self-self* turn pairs are shorter than time intervals in *other-self* turn pairs, which is intuitively clear because the speaker in *self-self* pairs does not have to process the message first in contrast to *other-self* pairs. This information will be integrated into the concept of user's **interaction profile** inspired by [8], which will be used for highly-adaptive user models.

We build an interaction profile for each user based on the following observable parameters: time stamp, turn length, turn and ancestor features (e.g. *self* or *other*, FPP or SPP, lexical or non-lexical) and type of action (e.g. greetings, repairs and disagreements). Since it is not known at the beginning, if the user is a slow-typer, which device she uses, whether she tends to produce a sequence of short messages (like L in Fig. 2) or one longer turn, the first session or the first minutes of the first chat can be used to build a knowledge base for responsiveness prediction. Our design aims are prediction of timing in user's activity, recognition of the end of increments, agent's reaction on missing response and accident handling. The agent does not have to simulate increments, but it should be able to recognise them.

Contribution

We proposed a corpus-driven approach for responsiveness analysis in IM interaction. We argue that responsiveness values combined with turn features can be effectively used for creation of interaction profiles which will significantly improve user models in HCI, especially in chatbot design, making timing in interaction more natural.

References

[1] D. Avrahami and S. Hudson. Responsiveness in instant messaging: Predictive models supporting inter-personal communication. In *HCI*, pages 731–740, 2006.

[2] S. Danilava, S. Busemann, C. Schommer, and G. Ziegler. Towards computational models for a long-term interaction with an artificial conversational companion. In *ICAART'13*, 2013. To appear.

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

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

[5] Chatbots directory. http://www.chatbots.org/.

[6] A. Raux and M. Eskenazi. Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. In *SIGdial '08*, pages 1–10. ACL, 2008.

[7] E. A. Schegloff. *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Cambridge University Press, 1 edition, 2007.

[8] T. Spranz-Fogasy. *Interaktionsprofile: Die Herausbildung individueller Handlungstypik in Gesprächen.* Radolfzell: Verlag für Gesprächsforschung, 2002.