Laughter in Human–Agent Interaction: Capturing the Temporal Relationship Between the Laughable and Laughter

Talha Bedir[1], Vanessa Vanzan[1], Amelie Robrecht-Hilbig[1],

Mathias Barthel[2,3], Erik Lagerstedt[1], Christine Howes[1]

[1]University of Gothenburg [2]University of Cologne [3]Leibniz Institute for the German Language

1 Introduction

1.1 Laughter in human-agent interaction

Laughter is an important interactional resource that contributes to the sequential organization of conversation, the management of social relationships, and the negotiation of meaning (Glenn, 2003; Jefferson, 1984; Clark, 1996). Although often associated with humour, laughter serves diverse communicative functions that depend on interactional context and the laughable (Provine, 2001; Mazzocconi et al., 2022; Ginzburg et al., 2020; Koutsombogera and Vogel, 2022). As conversational agents become increasingly integrated into everyday life, studying laughter in human–agent interaction (HAI) requires annotation schemes that account for both the laughable and interactional timing. Our analysis is based on the Barthel corpus (Barthel et al., 2023), comprising recordings of six German households interacting with Alexa.

1.2 Existing annotation frameworks

Existing annotation frameworks describe laughter in terms of its pragmatic function, interactional role, and relationship to the laughable (Mazzocconi et al., 2022; Koutsombogera and Vogel, 2022; Ginzburg et al., 2020). In HAI, laughter has been classified by triggers including TTS errors, discourse atypicality, timing errors, humorous statements, conversation endings, and post-own-speech laughter (Booker et al., 2024). While this framework extends laughter annotation, applying it to our corpus revealed cases that were difficult to annotate, particularly in cases where the laughable depended on or consisted of the temporal (mis-)organization of the interaction.

1.3 Timing as an annotation challenge

Temporal coordination is a fundamental aspect of spoken interaction (Blohm and Barthel, 2025; Sacks et al., 1974; Levinson and Torreira, 2015). Turn-taking differs between human-human and human-agent interaction, with users exhibiting longer response gaps and different turn-taking behaviour when interacting with artificial agents than with other humans (Barthel, 2025; Johansson et al., 2014; Skantze, 2021).

However, it remains unknown whether and how this difference affects laughter timing. Does it take us longer to start laughing when interacting with an artificial agent? And is laughter in human-agent interactions systematically timed in the same ways as in human-human interactions?

1.4 Present work

As part of our ongoing work on laughter in human-agent interaction (Vanzan et al., 2026), three annotators annotated 12 representative laughter events (10% of the corpus). During this process, we encountered a challenge concerning the coding of the temporal relationship between the laughable and the laughter event: our way of coding timing of the laughter did not differentiate between laughables.

This paper addresses the following methodological question: *How should the temporal relationship between the laughable and the laughter event be represented during annotation?* We focus on the temporal component of the annotation scheme and discuss its implications for laughter annotation in HAI. The Annotation scheme is provided in Appendix A.

2 Examples and discussion

While annotating the Alexa data, we encountered a number of sequences which required different timing schemes due to the different laughable objects. In this section, we will present the four main categories of laughter-containing sequences of user-Alexa interactions that we identified in our data.

Agent-response laughable. The most basic laughter exchange with Alexa follows a prompt-response-laughter sequence. In the example shown in Figure 1, one of two co-present users asks the question, "Alexa, what is the most popular sex position?"[1] and after a long pause Alexa responds "Unfortunately I don’t know", which triggers laughter by both users.

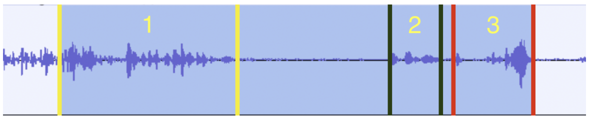

Figure 1: Example of a sequence including an Agent-response laughable. 1: [Adult 1] "Alexa what is your favorite sex position?", 2: [Alexa] "Unfortunately I do not know", 3: [Adult 1, Adult 2] laugh.

Human-utterance laughable. Another relatively straightforward category of cases is one speaker laughing at something that another speaker said in response to or in conversation with Alexa. In Figure 2, the laughter is triggered by the prompter’s utterance following Alexa’s response.

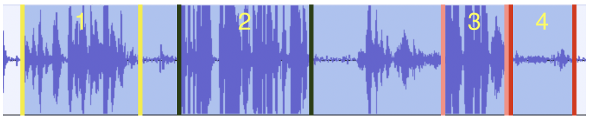

Figure 2: Human-utterance laughable. 1: [Child 1] "Alexa you are the best woman in the world. Can we marry each other?" 2: [Alexa] "According to Wikipedia: The marriageable age in Germany is 18 years." 3: [Child 1] "but I’m already eighteen". 4: [Child 2] laughs.

While these first two cases might make it seem straightforward to simply annotate the endpoint of the last turn as the moment the laughable becomes interactionally effective, a lot of the laughter instances in our collection actually overlap with Alexa’s response. In such cases, categories based solely on the laughable’s endpoint (Booker et al., 2024) do not capture when the laughable becomes relevant. Consider the following example containing a laughable that is not a speaking turn proper.

Agent-output laughable. These instances occur especially if the laughable is not the entire response but a part of it. For example, the long media output that is shown in Figure 3 is a song automatically sung by Alexa after the user prompt "I love you." It triggers multiple laughter episodes before the song ends.

[1]All examples were translated from German into English by the authors. For more information about the corpus, see (Barthel et al., 2023; Vanzan et al., 2026).

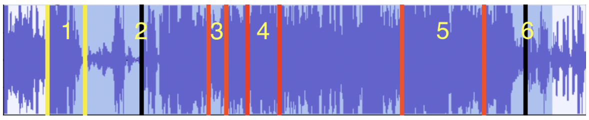

Figure 3: Agent-output laughable. Interaction: Alexa reacts to the prompt "Alexa, I love you" with a pre-installed song. 1: [Child 1] "Alexa, I love you", 2-6 [Alexa] plays a pre-installed song, 3: [Child 1] laughs, 4: [Parent] laughs, 5: [Child 1, Child 2, Parent] laugh.

Within-response laughable.The phenomenon of user laughter overlapping with Alexa’s response is not unique to media output such as the song prompted by "I love you" in Alexa devices–it is observed in Alexa’s regular answers as well. In the example in Figure 4, Alexa starts a long response containing several laughables that trigger multiple laughter episodes during the response, which makes temporal annotation necessary to identify the exact laughable.

Figure 4: Within-response laughable. 1: [Adult 1] "Alexa, who is Donald Duck?". 2-5 [Alexa] answers with a long explanatory text, 3: [Adult 1] laughs, 4: [Adult 2] laughs.

The examples in Figures 3 and 4 make it untenable to annotate the endpoint of Alexa’s response as the moment when laughter becomes relevant, since the laughable is not Alexa’s entire response but rather specific parts of it. These examples show that identifying the laughable requires temporal annotation.

Limitations and outlook

The annotation scheme and taxonomy are still being refined as additional interactional patterns emerge from the corpus. Consequently, some laughter events so far remain difficult to classify consistently, and the current categories may not yet capture the full range of temporal relationships between the laughable and laughter onset. Capturing when it emerges, rather than the boundaries of the agent’s response, supports consistent annotation and future analyses of laughter timing in HAI. Future work will include a formal inter-annotator agreement study to evaluate the reliability of the proposed annotation scheme and to identify categories requiring further refinement. Once annotation has been completed, the resulting corpus will enable quantitative analyses of the temporal organization of laughter in human–agent interaction, including the relative frequencies of different laughables and their temporal relationship to laughter onset. The annotation scheme will also make it possible to investigate how interactional factors such as participant configuration, co-presence, and users' age groups relate to laughter production.

Acknowledgments

This research was supported by ERC Starting Grant DivCon: Divergence and convergence in dialogue: The dynamic management of mismatches (101077927). Howes was further supported by the Swedish Research Council grant (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

References

Mathias Barthel. 2025. Alexa, you are too slow! invariant turn-transition times and conversational flow in natural human – voice agent interaction. In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, pages 135–145.

Mathias Barthel, Henrike Helmer, and Silke Reineke. 2023. *First Users' Interactions with Voice-Controlled Virtual Assistants: A Micro-Longitudinal Corpus Study*. University of Maribor, Maribor, Slovenia.

Stefan Blohm and Mathias Barthel. 2025. Why so cold and distant? effects of inter-turn gap durations on observers' attributions of interpersonal stance. In *Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, pages 125–134.

Nynaeve Perkins Booker, Michelle Cohn, and Georgia Zellou. 2024. Linguistic patterning of laughter in human-socialbot interactions. *Frontiers in Communication*, 9:1346738.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

Jonathan Ginzburg, Chiara Mazzocconi, and Ye Tian. 2020. Laughter as language. *Glossa: A Journal of General Linguistics*, 5(1):1–51.

Phillip Glenn. 2003. *Laughter in Interaction*. Cambridge University Press, Cambridge.

Gail Jefferson. 1984. On the organization of laughter in talk about troubles. In *Structures of Social Action: Studies in Conversation Analysis*, pages 346–369.

Martin Johansson, Gabriel Skantze, and Joakim Gustafson. 2014. Comparison of human-human and human-robot turn-taking behaviour in multiparty situated interaction. In *Proceedings of the 2014 Workshop on Understanding and Modeling Multiparty, Multimodal Interactions (UM3I)*, pages 21–26, Istanbul, Turkey. Association for Computing Machinery.

Maria Koutsombogera and Carl Vogel. 2022. Understanding laughter in dialog. *Cognitive Computation*, pages 1–16.

Stephen C. Levinson and Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6:731.

Chiara Mazzocconi, Ye Tian, and Jonathan Ginzburg. 2022. What's your laughter doing there? a taxonomy of the pragmatic functions of laughter. *IEEE Transactions on Affective Computing*, 13(3):1302–1321.

Robert R. Provine. 2001. *Laughter: A Scientific Investigation*. Penguin.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: A review. *Computer Speech & Language*, 67:101178.

Vanessa Vanzan, Talha Bedir, Vladislav Maraev, Erik Lagerstedt, Mathias Barthel, and Christine Howes. 2026. Fart gags and prudish machines: Laughter in human-agent interactions. In *Proceedings of the 13th International Conference on Human-Agent Interaction (HAI '25)*, pages 265–273, New York, NY, USA. Association for Computing Machinery.

A Annotation scheme

Each laughter event was annotated independently using the annotation dimensions described below.

Laughter onset Timestamp indicating the onset of the laughter event, recorded in minutes and milliseconds (e.g., `2:13122`).

Alexa onset Timestamp indicating the onset of Alexa's response (e.g., `0:12423`).

Alexa offset Timestamp indicating the end of Alexa's response (e.g., `0:17523`).

Laughable The perceived target of the laughter.

- `human` – a human or their behaviour;
- `agent speech` – Alexa's spoken response;
- `agent output` – Alexa's non-speech output (e.g., music or sound effects);

- `interaction` – the interaction itself;
- `unknown` – the target cannot be reliably inferred.

Whenever the available evidence was insufficient, annotators selected `unknown`.

Intensity Perceived intensity of the laughter, independent of its duration.

- `exhale` – barely voiced or breathy laughter;
- `chuckle` – light chuckles or giggles;
- `full` – clearly voiced laughter;
- `other` – laughter not fitting the above categories.

Laughter accompanied by speech Whether the laughter overlaps with speech.

- `yes` – laughter is produced while speaking;
- `no` – laughter occurs without simultaneous speech.

Laugher Identity of the producer of the laughter.

- `user` – the user interacting with the voice assistant;
- `other` – another person present;
- `VA` – the voice assistant.

Laugher age group Estimated age group of the person producing the laughter.

- `child`;
- `adult`.

Interactant age group Estimated age group of the person interacting with the voice assistant.

- `child`;
- `adult`.

Co-presence Number of other audible people present during the laughter event.

- `0` – no one else present;
- integer value – number of co-present people;
- `unknown`.

Relationship Relationship between the laugher and each co-present person (multiple values permitted).

- `parent-child`;
- `child-parent`;
- `sibling`;
- `partner`;
- `friend`.

Multiple values were separated by commas (e.g., `child-parent, sibling`).

Other's role Role of each co-present person in the interaction.

- `involved` – actively participating in the interaction;
- `bystander` – present but not participating;
- `N/A` – no other person present.

When multiple people were present, values were separated by commas.

Local Whether the laughable can be inferred from the local interaction alone.

- `yes`;
- `no` – broader interactional context is required.

Phase of usage Free-text description of the broader interactional episode (e.g., requesting jokes, requesting music, testing Alexa's capabilities).

Annotator Identifier(s) of the annotator(s).

Comments Optional free-text comments.