Conveying the Perception of Humor Arising from Ambiguous Grammatical Constructs in Human-Chatbot Interaction

Juan Carlos Farah
juancarlos.farah@epfl.ch
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

Vandit Sharma
vanditsharma@iitkgp.ac.in
Indian Institute of Technology Kharagpur
Kharagpur, India

Sandy Ingram
sandy.ingram@hefr.ch
College of Engineering and Architecture of Fribourg
Fribourg, Switzerland

Denis Gillet
denis.gillet@epfl.ch
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

ABSTRACT

Chatbots have long been advocated for computer-assisted language learning systems to support learners with conversational practice. A particular challenge in such systems is explaining mistakes stemming from ambiguous grammatical constructs. Misplaced modifiers, for instance, do not make sentences ungrammatical, but introduce ambiguity through the misplacement of an adverb or prepositional phrase. In certain cases, the ambiguity gives rise to humor, which can serve to illustrate the mistake itself. We conducted an online experiment with 400 native English speakers to explore the use of a chatbot to harness such humor. In an interaction resembling an advanced grammar exercise, the chatbot presented participants with a phrase containing a misplaced modifier, explained the ambiguity in the phrase, acknowledged (or ignored) the humor that the ambiguity gave rise to, and suggested a correction. Participants then completed a questionnaire, rating the chatbot with respect to ten traits. A quantitative analysis showed a significant increase in how participants rated the chatbot's personality, humor, and friendliness when it acknowledged the humor arising from the misplaced modifier. This effect was observed whether the acknowledgment was conveyed using verbal, nonverbal (emoji), or mixed cues.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → **Computer-assisted instruction**.

KEYWORDS

chatbots, humor, language learning, misplaced modifiers, emoji

ACM Reference Format:
Juan Carlos Farah, Vandit Sharma, Sandy Ingram, and Denis Gillet. 2021. Conveying the Perception of Humor Arising from Ambiguous Grammatical Constructs in Human-Chatbot Interaction. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21), November 9–11, 2021, Virtual Event, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3472307.3484677

HAI '21, November 9–11, 2021, Virtual Event, Japan

ACM ISBN 978-1-4503-8620-3/21/11.
https://doi.org/10.1145/3472307.3484677

1 INTRODUCTION

Conversational agents, commonly known as *chatbots*, are programs that can interact using natural language [2]. Conceptualized by Alan Turing [49], chatbots have evolved from early keyword matching implementations (e.g., ELIZA [51]), to pattern matching agents built using the Artificial Intelligence Markup Language (e.g., ALICE [50]), to complex knowledge-based models trained on corpora comprising billions of words (e.g., Google's Meena [4]). Nowadays, chatbots come in a variety of architectures [11] and permeate almost every domain of human-computer interaction (HCI) [14, 30, 48].

In education, conversational interfaces were incorporated into learning environments as early as the 1970s [28]. Although their suitability for computer-assisted language learning (CALL) was quickly noted [38], it was only in the early 2000s that web-based systems drew wider attention to the potential such interfaces had for language practice [16, 21, 26]. A challenge for CALL systems, however, is to provide meaningful feedback for mistakes caused by ambiguous grammatical constructs [6, 35]. Sentences with misplaced modifiers, for instance, are not necessarily ungrammatical but can portray the wrong message due to the awkward placement of an adverb or prepositional phrase [31]. In some cases, misplaced modifiers can cause unintended humor, as exemplified in: "*My client has discussed your proposal to fill the drainage ditch with his associates*" [20]. If chatbots are expected to act as conversational partners, recognizing and handling these ambiguities appropriately is not only important for pedagogical purposes, but also from the perspective of human-agent interaction (HAI). Indeed, the Computers Are Social Actors (CASA) framework postulates that interactive computers are perceived as agents subject to the social norms of human interaction [37]. This perception becomes even more pronounced when human-like traits such as humor are conveyed by the computer [34]. Guided by CASA, our objective was to explore how a chatbot's acknowledgment of humor affected the way it was rated along ten different anthropomorphic traits. Moreover, to contextualize our study, we embedded our chatbot within a simulated language learning exercise. In the following section, we motivate our approach by presenting a literature review covering the intersection of humor, language learning, and chatbots.

2 BACKGROUND AND MOTIVATION

Several studies have emphasized the positive effects of using humor in the language classroom [1, 7, 15, 44, 47]. Bell [9] provided

suggestions for incorporating humor in second language (L2) learning, while Wu [53] implemented humor in English as a foreign language (EFL) learning and found improvements in students' English proficiency. The potential to use humor to explain ambiguous grammatical constructs was captured by Holmes [24], who paired misplaced modifiers with a depiction of their alternative meanings, thus illustrating the ambiguity they introduce. Nevertheless, the use of chatbots to harness humor arising from such grammatical constructs remains unexplored.

To inform the study of this potential new use case, we built on the design of chatbots equipped with humor capabilities [17, 45], as well as their application to language learning [10, 13]. Early implementations of chatbots for L2—such as CSIEC [26]—paved the way for research incorporating them into education with a focus on learner autonomy [3] and open learner models [29]. However, despite increased adoption of chatbots in CALL, researchers have identified several challenges, some of which (e.g., restrictive scripts, dependency on correct user input, inability to stay relevant) have been noted in recent work [22, 41]. Challenges surfaced by two studies in particular motivate our focus on ambiguous grammatical constructs. A first study [12] evaluated the language resources of chatbots to better understand their potential as a tool for teaching English as a second language (ESL), finding several problematic areas, such as their ability to handle grammatically incorrect input sentences and to incorporate their knowledge in follow-up questions. A subsequent study [13] investigated the linguistic accuracy of ESL chatbots, concluding that current chatbots did not make robust conversational partners, in part due to improper handling of nonsensical and meaningless, yet grammatically correct sentences.

The challenges highlighted by the aforementioned studies stress the fact that our understanding of the factors affecting chatbot user experiences is limited [19]. Our study contributes to addressing this gap in the research. Given that users are increasingly expecting humor in their interactions with chatbots [33], our focus was to shed more light on how a chatbot's use of humor affects user experience and to explore how these perceptions might be dependent on the nature of the humor and on the context in which the chatbot is deployed. By harnessing humor arising from misplaced modifiers and embedding the interactions within a task-oriented grammar exercise, we investigated chatbot user experiences that have not been addressed as of yet. Specifically, our approach was to examine whether a chatbot that reacted to humorous misplaced modifiers within an interaction resembling an advanced grammar exercise was rated more positively than one that completed the task without explicit acknowledgment of the humor. To assess this effect, we conducted an experiment evaluating the use of verbal and nonverbal cues to convey the perception of this type of humor. We present our experimental design in the following section.

3 METHODOLOGY

We followed a between-subjects, 2×4 factorial design, with two phrases and four acknowledgments. The phrases were selected from the results of a preliminary survey in which subjects were asked to rate the funniness of 45 sentences with misplaced modifiers. We chose the phrases "*The candidate was falsely accused of covering up a crime by the media.*" (P1) and "*Each rancher must choose the appropriate method of castration for himself.*" (P2), as they obtained the lowest and highest mean rating, respectively. The acknowledgments comprised a control, i.e., no acknowledgment (C), and three treatments, emoji-only (T1), verbal-only (T2), and verbal and emoji (T3). Hereafter, we refer to conditions by the treatment/phrase code combination (e.g., CP1, T3P2).

We used Prolific [40] to recruit native English speakers who were born and currently live in the United States. For each of the eight conditions, we recruited 50 participants. A Kruskal-Wallis H test on the results of a pre-questionnaire indicated that there were no significant differences in participants' familiarity with chatbots ($H = 2.89$, $p = 0.895$) or frequency of interactions with chatbots ($H = 1.56$, $p = 0.980$). However, chi-square tests showed slight discrepancy in gender balance for conditions T1P2 ($\chi^2 = 5.80$, $p = 0.016$) and T3P1 ($\chi^2 = 3.92$, $p = 0.048$), while ANOVA tests indicated a slight deviation from the global mean age ($\bar{x} = 32.00$, $SD = 11.75$) in conditions CP1 ($\bar{x} = 36.72$, $SD = 13.14$) and T1P1 ($\bar{x} = 27.1$, $SD = 10.38$).

For all conditions, we configured a rule-based chatbot interface to run three short exchanges with participants. The first exchange introduced the grammatical context of the interaction. The chatbot explained that it had been reviewing an article and had found a sentence that stood out in the text. It then presented either P1 or P2 and asked the participant if the meaning of the sentence was clear. The participant could reply with "Yes", "No", or "I'm not sure". The second exchange consisted of an explanation of the ambiguity caused by the misplaced modifier, with variations across the experimental conditions. In the control condition, the exchange consisted solely of the explanation. In the treatment conditions, the explanation was followed by (i) a *rolling on the floor laughing* (ROFL) emoji (T1), (ii) the text "I find it hilarious!" (T2), or (iii) both the text and the emoji (T3). In all conditions, the exchange ended with a question asking participants if they agreed with the interpretation. Once more, participants could reply with "Yes", "No", or "I'm not sure". Finally, the chatbot proposed a reworded version of the original phrase to address the misplaced modifier. It then asked participants if the suggestion was good. As before, participants could reply with "Yes", "No", or "I'm not sure". A sample interaction for the T1P2 condition is depicted in Fig. 1.

Participants then completed a questionnaire in which they rated the chatbot. To select the traits along which participants rated the chatbot, we first adapted questions addressing the perception of social traits (e.g., likability, friendliness, attentiveness) on artificial agents [8, 54]. Second, we incorporated items to address conversation-specific traits such as naturalness and continuity [23]. Third, we included questions regarding personality and humor. Finally, we included a question on grammar, which has been identified as a barrier for the adoption of chatbots as conversational partners [13]. The result was a post-questionnaire with two main sections. In the first section, we asked participants to rate the chatbot using a five-point Likert scale (1 = "least", 5 = "most") with respect to seven traits. In the second section, we asked participants to respond to three statements concerning three different traits using a five-point Likert scale (1 = "strongly disagree", 5 = "strongly agree"). Within each section, questions were randomly shuffled for each participant. In summary, the ten traits captured by the instrument were (i) personality, (ii) humor, (iii) common sense, (iv)

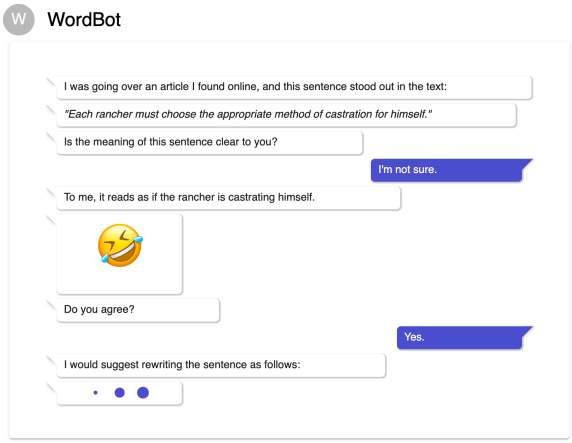

Figure 1: A sample interaction with the chatbot.

friendliness, (v) likability, (vi) naturalness, (vii) continuity, (viii) rapport, (ix) attentiveness and (x) grammar.

4 RESULTS

For all ten traits considered, the results of Mann-Whitney U tests and the differences in mean Likert score between treatment and control conditions are summarized in Table 1. Each trait was considered independently, with a significant result indicating that a treatment resulted in a significantly different distribution of ratings from its respective control. In this section, we report on the results of interest. When included, an increase (+) or a decrease (-) in mean Likert score is with respect to the corresponding control condition. For brevity, all tests refer to Mann-Whitney U tests.

Both **personality** and **humor** showed highly statistically significant increases across all treatment and phrase combinations. For personality, the largest increase for P1 (+1.02) was seen in T3 ($U = 1856.0$, $p < 0.001$), while for P2 it was (+1.10) in T1 ($U = 1940.5$, $p < 0.001$). The smallest increase for P1 (+0.78) was in T2 ($U = 1693.5$, $p = 0.002$), while for P2 (+0.94), it was in T3 ($U = 1861.5$, $p < 0.001$). Pairwise tests between treatment conditions showed no significant differences. For humor, the highest increase was seen in T1, both for P1 (+1.94, $U = 2199.5$, $p < 0.001$) and P2 (+1.96, $U = 2202.0$, $p < 0.001$). The smallest increase was seen in T3, both for P1 (+1.84, $U = 2152.0$, $p < 0.001$) and P2 (+1.72, $U = 2127.5$, $p < 0.001$). Again, pairwise tests between treatment conditions showed no significant differences. An increase in **friendliness** was also seen across all treatment and phrase combinations, though it was not statistically significant for the T2P2 (+0.32, $U = 1476.5, p = 0.085$) and T3P2 conditions (+0.08, $U = 1289.5, p = 0.771$). Other treatment conditions had consistent increases ranging from +0.34 ($U = 1515.0, p = 0.046$) for T3P1 to +0.40 ($U = 1554.5, p = 0.019$) for T1P2. Pairwise tests showed a significant difference only between T1P2 and T3P2 ($U = 971.0, p = 0.031$). **Continuity**, **rapport**, **grammar**, **common sense**, and **likability** all showed mixed effects, with significant decreases observed in T2P1 for rapport (−0.36, $U = 981.0, p = 0.050$) and T3P2 for common sense (−0.48, $U = 894.5, p = 0.010$). Pairwise tests between the treatment conditions within each trait showed a significant difference between T3P2 and T2P2 for grammar ($U = 975.5, p = 0.040$), and T3P2 and T1P2 for common sense ($U = 769.0, p < 0.001$).

Finally, the results for **naturalness** and **attentiveness** showed a consistent decrease across all conditions except T3P1 for naturalness, which remained equal. Pairwise tests between treatment conditions for naturalness showed no significant differences. Of particular note is condition T3 for attentiveness, which experienced the largest decrease both for P1 (−0.30, $U = 1032.0$, $p = 0.100$) and P2 (−0.52, $U = 925.5$, $p = 0.019$). Pairwise tests between treatment conditions for attentiveness showed a significant difference only between T2P2 and T3P2 ($U = 931.0, p = 0.021$).

5 DISCUSSION

Although we observed statistically significant increases across all treatments for personality and humor, this was not the case for the other eight traits. However, there are a number of promising results and parallels with previous findings.

First, the increases for humor across all treatment conditions indicate that the direct effect of our manipulation—to acknowledge humor—was achieved. Moreover, the positive correlation between the effects on humor and personality is aligned with work that has highlighted a close relationship between both traits [33, 46]. The observed effects also have implications in practice. As personality has been found to help chatbots gain user trust and provide a more engaging and enjoyable experience [25], chatbots should incorporate behaviors that target improving how users perceive their personality. If this improvement can be achieved with a simple interaction similar to the ones used in this study, it would provide an alternative to the use of jokes, which have been widely used for introducing humor into HCI [17, 42, 43, 53], but might become repetitive if overused [19]. In language learning, one could envision a chatbot that sparsely incorporates the type of interaction used in this study, either by reacting to user inputs that contain misplaced modifiers or through short, predefined exercises on the topic.

Second, we did not observe any significant differences in naturalness, continuity, or grammar under any treatment condition. Since these are traits related to the conversational capability of our chatbot, our results suggest that the introduction of an acknowledgment did not affect the user's perception of the chatbot's ability to communicate. Ensuring naturalness and continuity of conversation when introducing diversions is especially important for task-oriented interactions, for which traditional views regarded humor as distracting and time-wasting [36]. Considering these conversational traits only, our results suggest that the interactions used in the treatment conditions pose little risk of eliciting significant unintended negative effects. These findings could be used to argue for implementing such interactions in task-oriented chatbots for CALL. Chatbots that can provide relatable explanations through humor without sacrificing user experience could be a positive enhancement for CALL systems. A recent study on the use of the writing software *Grammarly* suggested that it be used in conjunction with an academic learning advisor (ALA) to provide support due to occasionally confusing or inaccurate feedback [39]. Incorporating a chatbot that could serve this purpose when the ALA is not available, or in blended and remote learning contexts, could help mitigate

Table 1: Means and Standard Deviations (SD) for Five-Point Likert Scores by Trait and Phrase

Trait	Phrase	Control	Treatment		
		(C)	Emoji-Only (T1)	Verbal-Only (T2)	Verbal and Emoji (T3)
Naturalness	P1	4.04 (0.88)	3.72 (1.21)	4.02 (1.08)	4.04 (0.99)
	P2	4.14 (0.70)	4.12 (0.82)	4.04 (1.01)	3.82 (1.14)
Continuity	P1	4.28 (1.03)	4.34 (1.04)	4.22 (0.84)	4.16 (0.89)
	P2	4.20 (0.76)	4.26 (0.75)	4.08 (1.01)	4.00 (0.95)
Personality	P1	2.94 (1.17)	**3.78 (1.15)*****	**3.72 (1.11)****	**3.96 (1.09)*****
	P2	3.12 (1.06)	**4.22 (0.89)*****	**4.16 (0.91)*****	**4.06 (0.91)*****
Rapport	P1	4.18 (0.85)	4.02 (1.08)	**3.82 (0.96)***	3.84 (1.02)
	P2	3.88 (0.85)	4.06 (0.84)	4.02 (0.98)	3.74 (0.99)
Humor	P1	2.00 (1.09)	**3.86 (1.34)*****	**3.94 (1.08)*****	**3.84 (1.67)*****
	P2	2.50 (1.28)	**4.32 (0.87)*****	**4.46 (0.86)*****	**4.22 (0.84)*****
Grammar	P1	4.30 (0.79)	4.56 (0.61)	4.44 (0.67)	4.42 (0.86)
	P2	4.40 (0.76)	4.40 (0.70)	4.46 (0.68)	4.12 (0.85)
Common Sense	P1	4.06 (1.02)	4.10 (1.05)	3.92 (1.01)	4.12 (0.94)
	P2	4.18 (0.87)	4.34 (0.77)	4.00 (1.12)	**3.70 (0.95)****
Attentiveness	P1	4.44 (0.79)	4.24 (1.02)	4.18 (0.96)	4.14 (0.97)
	P2	4.18 (0.94)	4.12 (0.82)	4.16 (1.02)	**3.66 (1.15)***
Friendliness	P1	4.12 (0.92)	**4.48 (0.81)***	**4.48 (0.74)***	**4.46 (0.76)***
	P2	4.20 (0.90)	**4.60 (0.61)***	4.52 (0.61)	4.28 (0.78)
Likability	P1	4.20 (0.93)	4.22 (0.95)	4.30 (0.93)	4.18 (0.87)
	P2	4.18 (0.85)	4.24 (0.82)	4.26 (0.88)	3.98 (0.98)

Results in blue indicate an increase with respect to the control condition while those in red indicate a decrease. Results in **bold** are statistically significant ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

this issue. Nevertheless, significant decreases in the results of three other traits (rapport, common sense, and attentiveness)—though for each observed in only one condition/phrase combination—serve as a reminder that playful interactions such as the acknowledgment of humor can introduce an overall higher expectation of the chatbot, which the chatbot might not be able to extend to other traits [32].

6 LIMITATIONS AND FUTURE WORK

Our study has certain limitations that are worth addressing in future work. The first limitation stems from the use of only two phrases with possibly limited intrinsic funniness. Phrases could have also had other unmeasured attributes (e.g., offensive, confusing), introducing biases that were not captured in the preliminary survey. Further exploring the effect of the phrase on our results will provide more rigorous conclusions on the effects of the treatments. A second limitation concerns external validity. To minimize misapprehension, we only recruited native English speakers. Although misplaced modifiers also affect native speakers [18, 27], a logical next step would be to validate our current findings with ESL learners. Applicability to other languages and demographic groups also needs to be tested, as perceptions of human traits, especially humor, might be contingent on cultural and linguistic factors [5]. Moreover, our questionnaire was explicitly built to complement the short nature of the interaction with our chatbot, but traits such as humor and personality are multifaceted in nature and could be better captured by a more robust questionnaire. Another limitation is with respect to the interactions with our chatbot, which lacked real-world context, had a fixed script, and only allowed participants to reply using quick response buttons. Future research could benefit from embedding the interaction within a real-world setting and providing more flexibility, such as open-ended typing and conversational scripts that adapt to user inputs. Finally, while a disembodied, text-only conversational agent was suitable for our experiment, exploring whether the effects can be reproduced with embodied and/or voice-based agents would be useful in practice.

7 CONCLUSIONS

Our findings suggest that chatbots could indeed harness humor from ambiguous grammatical constructs to strengthen anthropomorphic perceptions of traits such as personality and friendliness. These findings contribute to the design of research-informed educational applications, and more specifically to the implementation of strategies for incorporating humor in human-chatbot interaction for language learning. Possible applications outside language learning include adaptations for introductory programming courses, where humor could be useful for teaching software engineering best practices given the wide use of chatbots on software development platforms [52]. In future work, we aim to consolidate the design of our chatbot to explore these extensions and replicate our study in a field experiment.

ACKNOWLEDGMENTS

This research has been co-funded by the European Union's Horizon 2020 program (grant agreement no. 781012).

REFERENCES

[1] Rima Aboudan. 2009. Laugh and Learn: Humor and Learning a Second Language. *International Journal of Arts and Sciences* 3, 3 (2009), 90–99.

[2] Bayan Abu Shawar and Eric Atwell. 2007. Chatbots: Are they Really Useful? *LDV-Forum* 22, 1 (2007), 29–49.

[3] Bayan Abu Shawar and Eric Atwell. 2007. Fostering Language Learner Autonomy Through Adaptive Conversation Tutors. In *Proceedings of the Fourth Corpus Linguistics Conference*. University of Birmingham, Birmingham, UK, 8 pages.

[4] Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. 2020. *Towards a Human-like Open-Domain Chatbot.* Technical Report. Google Research. arXiv:2001.09977 http://arxiv.org/abs/2001.09977

[5] Dana L. Alden, Wayne D. Hoyer, and Chol Lee. 1993. Identifying Global and Culture-Specific Dimensions of Humor in Advertising: A Multinational Analysis. *Journal of Marketing* 57, 2 (1993), 64–75. https://doi.org/10.1177/002224299305700205

[6] Luiz A. Amaral and Detmar Meurers. 2011. On Using Intelligent Computer-Assisted Language Learning in Real-Life Foreign Language Teaching and Learning. *ReCALL* 23, 1 (2011), 4–24.

[7] Lance Askildson. 2005. Effects of Humor in the Language Classroom: Humor as a Pedagogical Tool in Theory and Practice. *Journal of Second Language Acquisition and Teaching* 12 (2005), 45–61.

[8] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.

[9] Nancy D. Bell. 2009. Learning About and Through Humor in the Second Language Classroom. *Language Teaching Research* 13, 3 (2009), 241–258. https://doi.org/10.1177/1362168809104697

[10] Serge Bibauw, Thomas François, and Piet Desmet. 2019. Discussing with a Computer to Practice a Foreign Language: Research Synthesis and Conceptual Framework of Dialogue-Based CALL. *Computer Assisted Language Learning* 32, 8 (2019), 827–877. https://doi.org/10.1080/09588221.2018.1535508

[11] Jack Cahn. 2017. *CHATBOT: Architecture, Design, & Development.* Technical Report. University of Pennsylvania, Philadelphia, PA, USA.

[12] David Coniam. 2008. Evaluating the Language Resources of Chatbots for their Potential in English as a Second Language. *ReCALL* 20, 1 (2008), 98–116. https://doi.org/10.1017/S0958344008000815

[13] David Coniam. 2014. The Linguistic Accuracy of Chatbots: Usability from an ESL Perspective. *Text & Talk* 34, 5 (2014), 545–567. https://doi.org/10.1515/text-2014-0018

[14] Lei Cui, Shaohan Huang, Furu Wei, Chuanqi Tan, Chaoqun Duan, and Ming Zhou. 2017. SuperAgent: A Customer Service Chatbot for E-commerce Websites. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics - System Demonstrations* (Vancouver, BC, Canada). Association for Computational Linguistics, Stroudsburg, PA, USA, 97–102.

[15] Marc Deneire. 1995. Humor and Foreign Language Teaching. *HUMOR* 8, 3 (1995), 285 – 298. https://doi.org/10.1515/humr.1995.8.3.285

[16] O. V. Deryugina. 2010. Chatterbots. *Scientific and Technical Information Processing* 37, 2 (2010), 143–147. https://doi.org/10.3103/S0147688210020097

[17] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, and Kenji Araki. 2009. Activating Humans with Humor–A Dialogue System That Users Want to Interact With. *IEICE Transactions on Information and Systems* E92-D, 12 (2009), 2394–2401. https://doi.org/10.1587/transinf.E92.D.2394

[18] John R. Edlund and Olga Griswold. 2012. Non-Native Speakers of English. In *Concepts in Composition: Theory and Practice of the Teaching of Writing* (2nd ed.), Irene L. Clark (Ed.). Routledge, New York, NY, USA, 317–338.

[19] Asbjørn Følstad and Petter Bae Brandtzaeg. 2020. Users' Experiences with Chatbots: Findings from a Questionnaire Study. *Quality and User Experience* 5 (2020), 14 pages.

[20] Tenielle Fordyce-Ruff. 2011. Laughing All the Way to Court: Avoiding the Humor and Headaches Created by Misplaced Modifiers. *The Advocate* 54, 11/12 (2011), 37–38.

[21] Luke Fryer and Rollo Carpenter. 2006. Bots as Language Learning Tools. *Language Learning & Technology* 10, 3 (2006), 8–14.

[22] Luke K. Fryer, Kaori Nakao, and Andrew Thompson. 2019. Chatbot Learning Partners: Connecting Learning Experiences, Interest and Competence. *Computers in Human Behavior* 93 (2019), 279–289.

[23] Ryuichiro Higashinaka, Toyomi Meguro, Hiroaki Sugiyama, Toshiro Makino, and Yoshihiro Matsuo. 2015. On the Difficulty of Improving Hand-Crafted Rules in Chat-Oriented Dialogue Systems. In *Proceedings of APSIPA Annual Summit and Conference 2015* (Hong Kong, China). IEEE, New York, NY, USA, 1014–1018.

[24] Laura Holmes. 2019. *Misplaced Modifiers: Using Humor to Teach Complicated Grammar Rules.* Institute for Excellence in Writing, Locust Grove, OK, USA.

[25] Mohit Jain, Pratyush Kumar, Ramachandra Kota, and Shwetak N. Patel. 2018. Evaluating and Informing the Design of Chatbots. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS '18)*. ACM, New York, NY, USA, 895–906. https://doi.org/10.1145/3196709.3196735

[26] Jiyou Jia. 2004. CSIEC (Computer Simulator in Educational Communication): A Virtual Context-Adaptive Chatting Partner for Foreign Language Learners. In *Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT '04)*. IEEE, New York, NY, USA, 690–692.

[27] Pieter H. Joubert and Silvia M. Rogers. 2015. Language Pitfalls: Native English Speakers. In *Strategic Scientific and Medical Writing*. Springer, Berlin, Germany, 25–37. https://doi.org/10.1007/978-3-662-48316-9_4

[28] Roy Kaplow, Samuel H. Desch, David O. Pettijohn, Melvin H. Rodman, and Franklin C. Smith. 1973. Illustrations of Conversational, Inquiry, Problem-Solving, and Questionnaire Type Interactions within the TICS System. In *Proceedings of the 7th Annual Princeton Conference on Information Sciences and Systems*. Princeton University, Princeton, NJ, USA, 389–393.

[29] Alice Kerly, Phil Hall, and Susan Bull. 2007. Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models. *Knowledge-Based Systems* 20, 2 (2007), 177–185. https://doi.org/10.1016/j.knosys.2006.11.014

[30] Liliana Laranjo, Adam G. Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y. S. Lau, and Enrico Coiera. 2018. Conversational Agents in Healthcare: A Systematic Review. *Journal of the American Medical Informatics Association* 25, 9 (2018), 1248–1258. https://doi.org/10.1093/jamia/ocy072

[31] Mark Lester and Larry Beason. 2018. *The McGraw-Hill Education Handbook of English Grammar and Usage* (3rd ed.). McGraw-Hill Education, New York, NY, USA.

[32] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, CA, USA) *(CHI '16)*. ACM, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[33] Indrani Medhi Thies, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'Neill. 2017. How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians. In *Human-Computer Interaction - INTERACT 2017*, Regina Bernhaupt, Girish Dalvi, Anirudha Joshi, Devanuj K. Balkrishan, Jacki O'Neill, and Marco Winckler (Eds.). Springer, Cham, Switzerland, 441–459.

[34] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers. *Journal of Consumer Research* 26, 4 (2000), 323–339. https://doi.org/10.1086/209566

[35] Luis Morgado da Costa, Francis Bond, and Xiaoling He. 2016. Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications* (Osaka, Japan). The COLING 2016 Organizing Committee, Osaka, Japan, 107–116.

[36] John Morkes, Hadyn K. Kernal, and Clifford Nass. 1999. Effects of Humor in Task-Oriented Human-Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory. *Human-Computer Interaction* 14, 4 (1999), 395–435.

[37] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers Are Social Actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. ACM, New York, NY, USA, 72–78. https://doi.org/10.1145/191666.191703

[38] Gerald E. Nelson, Jean Renard Ward, Samuel H. Desch, and Roy Kaplow. 1976. Two New Strategies for Computer-Assisted Language Instruction (CALI). *Foreign Language Annals* 9, 1 (1976), 28–37.

[39] Ruth O'Neill and Alex M. T. Russell. 2019. Stop! Grammar time: University Students' Perceptions of the Automated Feedback Program Grammarly. *Australasian Journal of Educational Technology* 35, 1 (2019), 42–56. https://doi.org/10.14742/ajet.3795

[40] Stefan Palan and Christian Schitter. 2018. Prolific.ac–A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance* 17 (2018), 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

[41] Silvia Pokrivcakova. 2019. Preparing Teachers for the Application of AI-Powered Technologies in Foreign Language Education. *Journal of Language and Cultural Education* 7, 3 (2019), 135–153. https://doi.org/10.2478/jolace-2019-0025

[42] Michal Ptaszynski, Pawel Dybala, Shinsuke Higuhi, Wenhan Shi, Rafal Rzepka, and Kenji Araki. 2010. Towards Socialized Machines: Emotions and Sense of Humour in Conversational Agents. In *Web Intelligence and Intelligent Agents*, Zeeshan-Ul-Hassan Usmani (Ed.). InTech, Rijeka, Croatia, Chapter 9, 173–205. https://doi.org/10.5772/8384

[43] Scott Schanke, Gordon Burtch, and Gautam Ray. 2020. Estimating the Impact of 'Humanizing' Customer Service Chatbots. https://doi.org/10.31235/osf.io/xud6z

[44] John Robert Schmitz. 2002. Humor as a Pedagogical Tool in Foreign Language and Translation Courses. *Humor* 15, 1 (2002), 89–114.

[45] Heung-yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to XiaoIce: Challenges and Opportunities with Social Chatbots. *Frontiers of Information Technology & Electronic Engineering* 19, 1 (2018), 10–26. https://doi.org/10.1631/FITEE.1700826

[46] James A. Thorson and F.C. Powell. 1993. Sense of Humor and Dimensions of Personality. *Journal of Clinical Psychology* 49, 6 (1993), 799–809.

[47] Susan Trachtenberg. 1979. Joke-Telling as a Tool in ESL. *TESOL Quarterly* 13, 1 (1979), 89–99. http://www.jstor.org/stable/3585979

[48] Jay Trivedi. 2019. Examining the Customer Experience of Using Banking Chatbots and Its Impact on Brand Love: The Moderating Role of Perceived Risk. *Journal of Internet Commerce* 18, 1 (2019), 91–111. https://doi.org/10.1080/15332861.2019.1567188

[49] Alan M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236 (1950), 433–460. https://doi.org/10.1093/mind/LIX.236.433

[50] Richard S. Wallace. 2009. The Anatomy of A.L.I.C.E. In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Robert Epstein, Gary Roberts, and Grace Beber (Eds.). Springer, Dordrecht, The Netherlands, 181–210. https://doi.org/10.1007/978-1-4020-6710-5_13

[51] Joseph Weizenbaum. 1966. ELIZA—A Computer Program For the Study of Natural Language Communication Between Man and Machine. *Commun. ACM* 9, 1 (1966), 36–45. https://doi.org/10.1145/365153.365168

[52] Mairieli Wessel, Bruno Mendes De Souza, Igor Steinmacher, Igor S. Wiese, Ivanilton Polato, Ana Paula Chaves, and Marco A. Gerosa. 2018. The Power of Bots: Characterizing and Understanding Bots in OSS Projects. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 19 pages.

[53] Siew-Rong Wu. 2008. Humor and Empathy: Developing Students' Empathy through Teaching Robots to Tell English Jokes. In *Proceedings of the 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning* (Banff, AB, Canada) *(DIGITEL '08)*. IEEE, New York, NY, USA, 213–214. https://doi.org/10.1109/DIGITEL.2008.27

[54] Ran Zhao, Oscar J. Romero, and Alex Rudnicky. 2018. SOGO: A Social Intelligent Negotiation Dialogue System. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) *(IVA '18)*. ACM, New York, NY, USA, 239–246. https://doi.org/10.1145/3267851.3267880