Proceedings of the Twenty-Ninth International
Florida Artificial Intelligence Research Society Conference

CATCH: Injecting "Contextual and Timely Conversational Humor" into Lifelike Avatars

Josiah Wong and Avelino J. Gonzalez
University of Central Florida
{josiah.w.ucf@knights.ucf.edu, gonzalez@eecs.ucf.edu}

Abstract

Humor is an integral part of human interaction. In the same way that humor aids human conversation in a variety of situations, humor can also enable avatars to communicate with humans in a human-like manner. We investigated the technical feasibility of incorporating humor into lifelike avatars by developing CATCH, a humor engine module that can inject contextually appropriate and well-timed humor into an avatar's neutral response. CATCH examines contextual cues to select the most relevant humor for the situation and it dynamically determines when humor is appropriate via a laughter feedback system. Our experimental results indicate that CATCH consistently picks contextually appropriate humor on a balanced humor frequency and that humor increases the perceived human-likeness of our avatar.

Introduction

Humor is an important part of being human. It is a mechanism by which humans bond, communicate more effectively, and stir emotions (Binsted et al. 2006). Humor is also very complex with ongoing research still seeking to understand and represent humor in its various forms.

A virtual person, called an avatar, is limited in its capacity to completely understand any human with whom it interacts due to inevitable failings in its natural language system (Binsted and Ritchie 1995). However, those failings can be made less annoying with the use of humor (Binsted and Ritchie 1995). In fact, humor is expected from avatars capable of showing emotion as noted in (Binsted et al. 2006) from (Cowie 2000). Applications for humorous avatars include improved virtual customer service agents (Binsted and Ritchie 1995), education (Hollister et al. 2013), and personal digital assistants (Binsted et al. 2006).

Intrigued by the prospect of humorous intelligent agents, we investigated the technical feasibility of integrating humor within an avatar in order to make it more human-like. Thus, we developed the *Contextual And Timely Conversational Humor* (CATCH) module to enable a conversational lifelike avatar to provide contextually appropriate and well-timed humorous responses. We first discuss the background of humor and prior work in computational humor. Next, we cover our implementation of humor in an avatar through CATCH. Then, we describe the experiment by which we tested the efficacy of CATCH in making our avatar appear more human-like followed by experimental results, plans for future work, and a conclusion.

Background

Humor has many benefits. Most generally, humor makes a person likeable (Cann, Calhoun, and Banks 1997). Humor can promote joy in others and enable them to view things in a different light (Binsted et al. 2006). It can even lessen the blow of a potentially harsh message (Binsted and Ritchie 1995). Furthermore, humor has non-conversational benefits such as aiding the creative process in product design (Kudrowitz 2010) and creating catchy advertisements and headlines (Binsted et al. 2006).

For the purposes of communication, the above benefits of humor can also apply to an avatar. Additionally, humor can help a human empathize with an avatar. According to the Mutual Vulnerability Theory of Laughter (Simon 2012), laughter is an acknowledgement of a common vulnerability between two parties. Thus, an avatar can emphasize a similarity it has with a human by using humor to make the human laugh. However, the delivery of humor is just as important as the humor content itself. If the human doesn't understand the joke or if the avatar doesn't say the punchline with the right pronunciation or emphasis, then the humor may fall flat and fail.

In the literature, there are two approaches to implementing humor in computer systems. The first approach is a humor construction approach which constructs humor

word by word. For example, the JAPE system (Binsted and Graeme 1994) generates a variety of puns and the HAHAcronym system (Stock 2005) outputs satirical acronyms. Other examples include the approach with ontological semantics based on the General Theory of Verbal Humor (Hempelmann, Raskin, and Triezenberg 2006) and the work by Oaks involving structural ambiguity potentials and the insertion of words into joke formulas (Oaks, 2012).

The second approach is a humor selection approach which selects pre-written humor sentences or samples to output to the human. Microsoft's Cortana follows such an approach by outputting pre-written humorous responses to certain questions (Debois 2014). For instance, one of Cortana's responses to "What is the best search engine?" is "Bing, bing, bing! We have a winner!" More generally, a method for catering humor to different age groups is discussed in (Kadri 2012) by selecting humorous sentences from caches for different motivational categories which have different appeals among the age groups.

Our method uses the second approach. The humor engine module we created models how humans tell jokes to each other. When someone tells us a joke that we particularly enjoy, we remember it until an opportune time comes along when we can repeat the joke to someone else. In the same way, CATCH "remembers" humor samples that have been inserted beforehand into its database and it only needs to select the most appropriate humor sample to inject at the proper time into the avatar's response.

Our method's novelty lies in our focus on implementing humor in a conversation. Humor in this context must augment and not detract nor distract from the conversation. Thus, humor must be relevant to the topics being discussed or else it risks changing the subject. Also, humor must be told often enough to make the conversation interesting, but not so often that it becomes the sole focus of the conversation. Thus, our work concentrates on the criteria of contextual appropriateness and timeliness in conversational humor between a human and a lifelike avatar.

Methodology

In order to introduce an element of humor into a conversational lifelike avatar, we created CATCH, a humor engine module that can be consulted by an avatar in order to convert a neutral response into a humorous one. CATCH determines when it is appropriate to inject humor and it also determines which humor samples are relevant for the current conversation. In the following subsections, we describe CATCH's interface, the types of humor used by CATCH, the algorithm for determining when and how often humor should be used, and the algorithm for injecting a relevant humor sample into the avatar's neutral response.

Interface

In our environment, a person talks to an avatar via spoken natural language or through a touch screen. The avatar replies via spoken natural language. We confined the human-avatar interaction to that of a question-answer forum where the human asks questions about a given topic and the avatar responds with factual answers. However, this system can become repetitive to the human. With the addition of CATCH, the avatar can vary the dynamic of the conversation by occasionally injecting humor as seen in the following sample conversations:

Human: How large is the Pterodactyl? (1)

Avatar (no humor): The Pterodactyl had a wingspan of up to five feet long. How else may I assist you? (2)

Avatar (humor): The Pterodactyl had a wingspan of up to five feet long. Hee hee. You'll love this one. What is the scariest dinosaur? (2)

Human: What? (3)

Avatar: The Terror-dactyl! (4)

Human: Ha, ha, ha! (5)

Avatar: I knew you'd like that. What else may I help you with? (6)

Figure 1 shows how the avatar interacts with the human and CATCH; the numbers in the diagram correspond to those in the sample conversation above. In step 1, the human asks the avatar a question and the avatar formulates a response. In step 2, the avatar sends its neutral response to CATCH. If CATCH does not attempt humor, the response remains unchanged when the avatar sends it to the user. Otherwise, it is altered to be humorous which initiates a humor script. If the human plays along in step 3 by requesting the punchline to a joke setup, then the avatar will solicit the punchline from CATCH in step 4 and deliver it to the human. The human may or may not laugh in response in step 5. In the same step, the avatar tells CATCH whether or not the human laughed. Then, CATCH provides the avatar with the proper acknowledgement to say to the human in step 6.

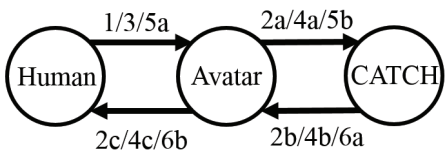

Figure 1: The avatar-CATCH interface in a sample conversation

Figure 2 outlines all possible scenarios of the humor script. The conversation in Figure 1 represents an instance where the humor script is carried from state 0 to states 1, 2, and then 3, resulting in success. Humor results in failure if

the human repeatedly refuses to request the punchline in state 1 or does not laugh at the avatar's humor in state 3. Dashed lines indicate human actions and solid lines indicate avatar actions. Retorts and witticisms are humor types without a setup and are explained in the next section.

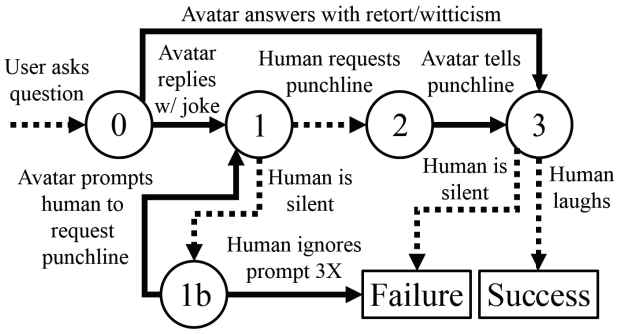

Figure 2: The humor script flow diagram if humor is attempted

Humor Types

The humor at CATCH's disposal was created and inserted by a human into CATCH's database. All humor is family-friendly in order to reduce the risk of offending someone.

CATCH uses three humor types. The first type is the *joke*, a humor sample that is composed of a setup and a corresponding punchline. A joke is inserted at the end of the avatar's neutral response. The second humor type is the *witticism*, a humor sample that is an afterthought and is appended to the end of the avatar's neutral response. Unlike the joke, a witticism is not meant to be meaningful if it is separated from the avatar's neutral response. The third humor type is the *retort*, a humor sample that directly addresses something the human has said and it precedes the avatar's neutral response.

> Joke: The Pterodactyl had a wingspan of up to five feet long. Hee hee. You'll love this one. What is the scariest dinosaur? … The Terror-dactyl!
>
> Witticism: The Pterodactyl had a wingspan of up to five feet long. That's why Pterodactyls get sore when they soar, see?
>
> Retort: When in doubt, the Pterodactyl is the guy to ask. Wait, he's extinct! … The Pterodactyl had a wingspan of up to five feet long.

Each humor sample is stored in CATCH's database as its own humor entry. Each humor entry is annotated with tags that assist CATCH in determining the humor sample's suitability for injection into a given neutral response by the avatar. The major tags are: associated knowledge base, humor type, associated topic, associated subtopic, humor content (as it should appear in subtitles), and humor content (as it should be verbally pronounced).

Humor Timeliness Algorithm

Humor is well-timed if the frequency at which it is attempted maximizes the human's satisfaction. Too few attempts will leave the human wanting more, but too many attempts will render the humor repetitive and boring. CATCH can change the timeliness of its humor during execution to suit the human's individual preferences.

The following pseudo-code for RELAY is called by the avatar and it shows *when* CATCH will attempt humor.

```
RELAY(Neutral_Response, Understood_Human)
1: Humor_Response := Neutral_Response
2: if Understood_Human is false then
3:       Humor_Response := HUH-RESPONSE( )
4:       return
5: if IS-SAFE(Neutral_Response) is false then
6:       return
7: Humor_Response := ANALYZE(Neutral_Response)
```

CATCH will only access its humor database if the conversation passes three conditions. The first condition states that the avatar must completely understand the human's question. If this check fails, CATCH randomly selects a generic witty comment designed for just this situation such as "I'm sorry. I couldn't hear you over the Beach Boys song playing in my head. Could you please try again or rephrase?" The second condition states that the conversation must not be about sensitive subjects such as violence or death. CATCH checks if any words from a list of "trigger words" associated with sensitive topics appear in the avatar's response. The third condition states that CATCH must have waited long enough since its last humor attempt.

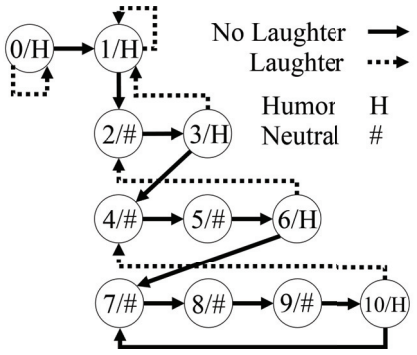

Figure 3: Finite state machine for humor timeliness algorithm

To check the third condition, CATCH uses a finite state machine (see Figure 3) which determines *how often* CATCH should attempt humor. CATCH begins in time state 0 and each row of states from top to bottom indicates a certain frequency of humor: always, frequently, occasionally, and rarely. If CATCH is in a *neutral* state, it will move to the next state in the row without attempting humor for the current avatar response. If CATCH is in a *humor*

state, CATCH will attempt humor for the current response. If the human laughs at its humor, CATCH will move to the first state in the next row up in the state machine and attempt humor one response sooner. Otherwise, it will move to the first state in the next row below and wait one additional response longer before attempting humor again.

CATCH uses a naïve laugh detection scheme that assumes that a human will respond to a humorous response exclusively with laughter or silence. The human's speech is converted into a string by the speech recognition system. However, human laughter does not consist of words and is unique to each person which complicates laugh recognition. Thus, CATCH assumes that *any* sound it hears while listening for laughter constitutes laughter, even if the human says something like, "That joke sucked!" A more complex laugh detection scheme is left for future research.

Contextual Appropriateness Algorithm

Once CATCH decides to attempt humor, CATCH selects a humor sample in three steps as seen in ANALYZE (which is called in RELAY). Step 1 determines the topic (e.g. Stegosaurus) and subtopic (e.g. diet) of the conversation by analyzing the avatar's neutral response. Step 2 uses this topic and subtopic to obtain the most relevant humor entry for injection. Step 3 alters the avatar's neutral response based on the chosen humor entry's humor type.

ANALYZE(Neutral_Response)
1: Humor_Response := Neutral_Response
2: Topics := GET-TOPICS(Neutral_Response)
3: Chosen_Entry := GET-ENTRY(Topics)
4: **return** INSERT-HUMOR(Chosen_Entry)

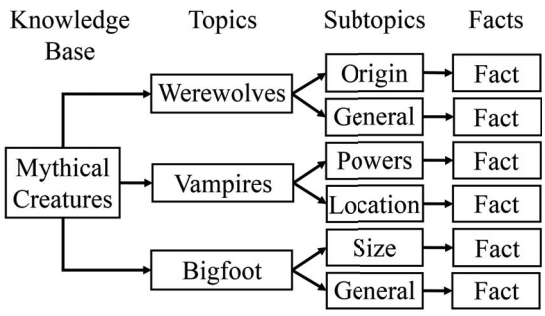

Figure 4: The Contextual Graph structure for an abbreviated version of the avatar's knowledge base

In the first function, GET-TOPICS, CATCH searches the avatar's neutral response for words that indicate conversational context. Our avatar's neutral response is previously derived from its knowledge base which is a Contextual Graph (see Figure 4). The avatar would have followed branches corresponding to words in the human's question through various contextual nodes until it arrived at the appropriate pre-written factual answer in a terminal action node. All words that are unique to a given topic or subtopic in the avatar's knowledge base are stored in CATCH's filter file. The topic and subtopic of the conversation are the topic and subtopic that have keywords in common with the avatar's neutral response. This topic and subtopic are correct to the extent that the avatar correctly understood the human.

GET-TOPICS(Neutral_Response)
1: Initialize Topic and Subtopic to null
2: Initialize all elements of Scores to 0
3: **for** every word *a* in Neutral_Response **do**
4: **for** every topic *t* in Possible_Topics **do**
5: **if** *a* is in *t*.Keywords **then**
6: Topic := *t*
7: **for** every subtopic *s* in Possible_Subtopics **do**
8: **if** *a* in *s*.Keywords **then**
9: Increment score for *s* in Scores
10: Subtopic := the subtopic with the highest score
11: **return** {Topic, Subtopic}

In the second function, GET-ENTRY, a list of humor entries sorted by relevance is obtained. The first humor entry that has not been previously used is returned. Then, the third function, INSERT-HUMOR, places the chosen humor entry into the avatar's neutral response. Jokes (preceded by a transition comment) and witticisms go at the end of the neutral response while retorts go before the neutral response (though in the case of a retort, the neutral response is delivered after laughter is acknowledged).

GET-ENTRY(Tags)
1: Best_Entries := FIND-ENTRIES(Tags)
2: Remove entries from Best_Entries that are in Blacklist
3: Store in Chosen_Entry the humor entry with highest relevance in Best_Entries, pick random to break tie
4: APPEND(Blacklist, Chosen_Entry)
5: **return** Chosen_Entry

To facilitate the calculation of a humor entry's relevance to a conversation, all humor entries have an associated knowledge base, topic, and subtopic. If a humor entry is relevant to all subtopics, its subtopic is the same as its topic. Likewise, the topic of a humor entry that works with all topics is the same as its knowledge base. Examples of the latter case may have "interchangeable" subjects. For example, the joke, "What weighs more than a dinosaur? Her shadow!" could be construed to fit multiple conversational contexts by replacing "dinosaur" with the name of the dinosaur currently being discussed. Between two valid humor entries, the one that is usable in fewer conversational contexts is considered more relevant.

Each humor entry calculates its own relevance given the conversational topic and subtopic per the specification in Figure 5 where a lower number in the relevance column indicates higher relevance. There is also a differentiation between multi-topic entries that can insert the topic as the subject and those that can't.

Relevance	1	2	3	4	5	6	7	8	9	10
Only this topic	X		X			X				
All topics (can insert)		X		X			X			
All topics (can't insert)					X			X	X	
Only this subtopic	X	X			X					
All subtopics			X	X				X		

Figure 5: Humor entry relevance specification

Assessment

In order to test the effectiveness of humor in increasing the human-likeness of an avatar, we had 44 college students interact with our humor-equipped avatar and then fill out a survey. Most of the volunteer test subjects were computer science and engineering majors. There were 35 men and 9 women. Each test subject was brought into the lab hosting our avatar, JAMES.

JAMES is an avatar that was previously developed in a collaboration between the University of Central Florida and the University of Chicago for an exhibit in the Orlando Science Center (Hollister et al. 2013). In our experiment, JAMES appeared on one computer screen and some possible questions that JAMES could answer appeared on an adjacent touch screen. JAMES used a Mythical Creatures knowledge base (e.g. vampires, werewolves, mermaids, etc.) for each test subject and "he" always began the conversation by introducing himself as a mythical creatures guru training to be a comedian.

After JAMES's self-introduction, the study facilitator spoke with JAMES in order to demonstrate to the test subject that one can 1) communicate with JAMES by speaking into the microphone or by touching a question on the touch screen and 2) choose whether or not to laugh into the microphone in response to a humorous reply by JAMES. After this, the test subject spent about 5 to 10 minutes asking JAMES questions about mythical creatures.

After talking with JAMES, the test subject filled out an anonymous survey asking for demographic information, open-ended suggestions for improvement, and his or her perceptions of the avatar. The following four questions were asked to assess the test subject's perceptions of the avatar's humor and human-likeness:

> Questions 1-3: Which of the following do you believe is the most accurate statement?
>
> Q1: "The humor was ___ to the context of the conversation." Choices: relevant, NOT relevant
>
> Q2: "The avatar injected humor into the conversation ___." Choices: too many times, too few times, the right number of times
>
> Q3: "Overall, the humor ___." Choices: was not funny at all, was 'cute', made me smile, made me laugh out loud
>
> Q4: What is your reaction to the statement, "The avatar's humor made the avatar seem more human-like"? Choices: "I strongly disagree." "I disagree." "I neither agree nor disagree." "I agree." "I strongly agree."

Results

As we can see from our experimental results in Figure 6, CATCH selected humor that was contextually appropriate to the conversation for about 95% of the 44 test subjects' conversations. This is by far CATCH's strongest feature. CATCH's timeliness was evenly balanced between using humor too many times and too few times. The variance of this metric would ideally be very low and the data indicate that this feature can be improved to reduce the variance.

Even though humor funniness was out of CATCH's control because all humor content was provided beforehand by a human, funniness was a possible lurking variable which may have affected test subjects' perceptions of our other metrics. Thus, it is encouraging to see that only three test subjects did not find CATCH's humor funny.

Finally, about 56.8% of the test subjects agreed or strongly agreed that humor increased the avatar's human-likeness. JAMES still has yet to pass the Turing Test and personal observations of the test subjects implied that humor was not completely successful at masking JAMES's shortcomings for other human-like traits. Thus, it is understandable that this percentage is not much higher than 50% because humor is only one small aspect of humor nature.

Some common suggestions for improvement from test subjects included 1) improve speech recognition, 2) use a less robotic voice (JAMES used a text-to-speech voice), 3) insert humor samples in the middle of the avatar's neutral response instead of just at the beginning or end of it, and 4) recognize smiles and quiet snickers that are not detectable with a microphone, instead of relying only on laughter to determine how a test subject received the avatar's humor.

There were a couple items that limit the generality of our results. First, the pool of test subjects was mostly composed of male engineers. Thus, our results could vary for

different demographic groups. Second, our experiment lacked a control group and the survey questions were subjective in nature. Thus, we cannot reject the null hypothesis regarding Q4. However, we wanted to see whether or not our implementation of humor in avatars was effective *at all.* The variability of demographic groups' perceptions was outside the scope of this research and a positive trend in the results is at the very least qualitatively evident.

Q1	Not Relevant	Relevant
	2 (4.5%)	42 (95.5%)

Q2	Too Few Times	Just Right	Too Many Times
	13 (29.5%)	21 (47.7%)	10 (22.7%)

Q3	Not Funny	Cute	Smile	LOL
	3 (6.8%)	15 (34.1%)	25 (56.8%)	1 (2.3%)

Q4	S. Disagree	Disagree	Neutral	Agree	S. Agree
	1 (2.3%)	9 (20.5%)	9 (20.5%)	24 (54.5%)	1 (2.3%)

Figure 6: The experimental results on the effects of CATCH

Future Work

The next CATCH-enabled system will seek to rectify the issues brought up by the test subjects. One test subject suggested that responses to CATCH's humor should be evaluated on a graduated scale instead of in a completely binary fashion (laughed/did not laugh). Other improvements may enable CATCH to 1) control the avatar's body and facial gestures, 2) learn which humor samples the human likes best and adapt accordingly, 3) respond to verbal feedback such as "Just give me the answer," or "That's great!", 4) learn jokes and humor from the human on the spot, 5) discern laughter from non-laughter, and 6) derive contextual clues from the user's question and not just the avatar's neutral response.

Additionally, we will also expand the scope of our investigation. Future surveys will include more questions to detect potential influences of humor funniness on test results. Furthermore, a sample of test subjects representing a greater demographic diversity will be obtained to measure the effectiveness and adaptability of the CATCH methodology for different genders, age groups, and cultures.

Conclusion

In this paper, we presented CATCH, a humor engine module which enables a conversational lifelike avatar to use humor when conversing with a human. CATCH injects humor into an avatar's neutral response to a human such that it is relevant to the conversational context and deployed at the right moments. In our experiment, we measured people's perceptions of CATCH's contextual appropriateness and humor timeliness, and the resultant effect of humor on the avatar's human-likeness. Our results indicate that CATCH consistently succeeds in selecting contextually appropriate humor and that overall, CATCH strikes the right balance between using humor too much and using it too little. Finally, a slight majority of test subjects agree that humor increases our avatar's human-likeness. We have thus shown that humor in conversational avatars is possible and worth further investigation.

References

Binsted, K.; Nijholt, A.; Stock, O.; Strapparava, C.; Ritchie, G.; Manurung, R.; and Waller, A. 2006. Computational Humor. *IEEE Intelligent Systems* 21(2): 59-69.

Binsted, K. and Ritchie, G. Using Humour to Make Natural Language Interfaces More Friendly. Unpublished. Ms. University of Edinburgh.

Binsted, K. and Graeme, R. 1994. An Implemented Model of Punning Riddles. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*. 633-638. Menlo Park, Calif.: AAAI Press.

Cann, A., Calhoun, L. G., and Banks, J. S. 1997. On the Role of Humor Appreciation in Interpersonal Attraction: It's No Joking Matter. *Humor: International Journal of Humor Research*, 10(1): 77-89.

Cowie, R. 2010. Describing the Emotional States Expressed in Speech. In Proceedings of the International Speech Communication Association Workshop Speech and Emotion, 11-18. New Castle, Northern Ireland, UK: International Speech Communication Association

Debois, D. 2014. Funny Things to Ask Cortana. *AmongTech*: http://www.amongtech.com/funny-things-to-ask-cortana/.

Hempelmann, C. F., Rask, V., & Triezenberg, K. E. (2006). Computer, Tell Me a Joke … but Please Make it Funny: Computational Humor with Ontological Semantics. *2006 AAAI FLAIRS Conference: Trends in Natural Language Processing.*

Hollister, J., Parker, S. T., Gonzalez, A. J., and DeMara, R. F. An Extended Turing Test: A Context Based Approach Designed to Educate Youth in Computing. In *Modeling and Using Context*, eds. Brezillion, P., Blackburn, P., and Dapoigny, R., 213-221. Berlin: Springer.

Kadri, F. L. 2012. Hansel and Gretel for All Ages: A Template for Recurring Humor Dialog. In *Proceedings of the 2012 AAAI Fall Symposium Series: Artificial Intelligence of Humor*, 26-29. Arlington, Virginia: AAAI Press.

Kudrowitz, B. M. 2010. HaHa and Aha! Creativity, Idea Generation, Improvisational Humor, and Product Design. Ph.D. dis., Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Oaks, D. D. 2012. On a Possible Generative Approach to Structurally Ambiguous Humor. In *Proceedings of the 2012 AAAI Fall Symposium Series: Artificial Intelligence of Humor*, 46-50. Arlington, Virginia: AAAI Press.

Simon, J. C. 2012. Computational Humor: Promises and Pitfals. In *Proceedings of the 2012 AAAI Fall Symposium Series: Artificial Intelligence of Humor*, 84-87. Arlington, Virginia: AAAI Press.

Stock, O. and Strapparava, C. 2005. HAHAcronym: A Computational System. In *Proceedings of the Association for Computational Linguistics*, 113-116. Ann Arbor, Michigan: ACL.