Harvard Data Science Review • Issue 3.2, Spring 2021

Computers Learning Humor Is No Joke

Thomas Winters[1]

[1]**Department of Computer Science, Science, Engineering & Technology Group, Katholieke Universiteit Leuven, Leuven, Belgium**

Published on: Apr 30, 2021

DOI: https://doi.org/10.1162/99608f92.f13a2337

Column Editor's Note: *Machine generation and comprehension of language is a difficult enough task, but it's not even funny how difficult it is for a machine to produce and detect humor. Data scientist Thomas Winters explains the reasons for this challenge, and addresses how much humor a computer can learn. (Spoiler: A bit).*

Keywords: computational humor, humor detection, humor generation

---

Can a computer have a sense of humor? On reflection, this question may seem paradoxical given that humor is such an intrinsically human trait. While some primates produce laughter (Gervais & Wilson, 2005; Preuschoft & Hooff, 1997), humans are the only known species that use humor for making others laugh. Furthermore, every known human civilization also has had at least some form of humor for making others laugh (Caron, 2002; Gervais & Wilson, 2005). Given this innately and intrinsically human skill, it might seem implausible for a machine to have any form of a sense of humor. In this article, we explore several computational techniques that have been used to detect and generate humor, and how these systems can help us improve our humor abilities and gain more insights into this uniquely human trait.

Human Sense of Humor

Over the whole of human history, many scholars investigated the inner workings of humor, resulting in an abundance of different theories for explaining verbally expressed humor. While no generally accepted all-encompassing theory of humor exists, most modern formal humor theories point to 'incongruity' as one of the important basic elements of humor (Gervais & Wilson, 2005; Hurley et al., 2011; Morreall, 1986). Incongruity denotes the incoherence between the interpretation of two parts of a joke, namely, its setup and punchline. One such popular incongruity theory is the incongruity-resolution (IR) theory, which is especially useful for computational models due to its mechanical (rather than descriptive) nature. The IR theory states that humor arises when an incongruity is resolved, meaning that one overarching, compatible interpretation for the whole text is discovered after hearing the punchline (Ritchie, 1999; Shultz, 1974; Suls, 1972). Others argued that incongruity and resolution are insufficient conditions for humor, stating that the interpretations must not replace one another and that the second interpretation must be of lower value than the first (Apter, 1982; Wyer & Collins, 1992). Another theory suggests that the two interpretations must overlap while also oppose one another to create humor (Raskin, 1984). As an example of IR, consider the following joke:

> Two fish are in a tank. Says one to the other:
> "You man the guns, I'll drive."

According to the IR theory, the setup creates an initial interpretation of two fish being in an aquarium. The punchline breaks this mental image, as aquaria generally lack weaponry and controllable mobility. Our brain thereupon searches for a reasonable explanation to fix this incongruity by considering previously heard text, and likely discovers the hidden interpretation that the fish were in a military vehicle all along. Resolving this

incongruity triggers our brain to release reward-related responses in its subcortical structures, causing the feeling of mirth and evoking laughter (Chan et al., 2012). The evolutionary reason for this behavior has been hypothesized as a reward for 'debugging' false presumptions and preventing them from polluting our mental world knowledge store (Hurley et al., 2011). While it is meant as an incentive to double-check candidate beliefs, humans have learned to exploit this behavior using humor, similar to how the advantageous desire for sweetness leads to our sugar addictions (Hurley et al., 2011).

Humor's frame-shifting prerequisite reveals its difficulty for a machine to acquire. It requires the right difficulty for the listener to not just trivially resolve the incongruity, while still ensuring the possibility of resolution. This substantial dependency on insight into human thought (e.g., memory recall, linguistic abilities for semantic integration, and world knowledge inferences) often made researchers conclude that humor is an 'AI-complete problem' (Attardo, 2001; Binsted et al., 2006; Hurley et al., 2011; Stock & Strapparava, 2006). This category denotes problems that are no less difficult than solving 'general intelligence' or 'strong AI.' In other words, genuine humor appreciation requires machines to have human-level intelligence, since it needs functionally equivalent cognitive abilities, sensory inputs, anticipation-generators and worldviews, as it otherwise must rely on error-prone heuristic approaches (Hurley et al., 2011).

Computational Humor Applications

Teaching computers humor tasks paves the way for various diverse practical applications. While not specifically designed for humor, recommendation engines (such as on YouTube and Netflix) already help users sift through humorous content to suit their preference using computational means (Davidson et al., 2010). Given that writers of all genres already use simple computational tools (such as online thesauri) for improving their writing, getting them to use smarter computational humor-driven interfaces for generating complex wordplay such as neologisms appears within reach (Gangal et al., 2017). Currently, comedy writing apps provide limited intelligent systems, with apps like The Gag[1] automatic estimating joke delivery time and Pitch[2] providing inspiring prompts. Similar to how Grammarly[3] improves users' general writing style for specific purposes, computational humor algorithms could potentially help augment our humor writing skills by performing automatic humor evaluation and providing suggestions or surprising associations (Gatti et al., 2015; Ritchie et al., 2007). Additionally, it could help translate wordplay, a notoriously difficult task (Chiaro, 2017; Vandaele, 2010).

With approximately 41% of virtual assistant users already viewing their virtual assistant as a friend (Kleinberg, 2018), computational humor could potentially further strengthen this bond. Since humor and creativity in language are an important trait in human friendships (Gray et al., 2015), the canned jokes told by virtual assistants do not suffice to emulate this aspect of friendships (Hempelmann, 2008; Lopatovska et al., 2020). Automatically tailoring generated humor to the user and producing jokes about shared experiences could potentially deepen these human-machine bonds (Binsted, 1995; Hempelmann, 2008), and might even be necessary for creating a generally intelligent system (Hurley et al., 2011).

To test formal humor theories, computational humor systems could provide a means for understanding the cognitive processes behind humor by producing humor according to a particular theory. Alternatively, by aggregating trends in large quantities of humor, humor detection algorithms can help linguists and cognitive scientists perform automated analyses to gain deeper insights into current humor trends. Both methodologies could help further the debate in search of a grand unifying theory of humor.

Generating Humor

Even though teaching computers the general concept of humor is a challenging undertaking, various researchers developed computational approaches both for humor detection and generation. By limiting the scope to specific joke types and by hand-coding patterns and features for them, researchers generally have been able to achieve several methods for detecting and generating humor.

JAPE (Joke Analysis and Production Engine) is one of the earliest large automated humor production systems (Binsted & Ritchie, 1994). It uses a hand-coded template-based approach to create punning riddles for children, such as:

What's green and bounces?
A spring cabbage.

What is the difference between leaves and a car?
One you brush and rake, the other you rush and brake.

It employs templates like "What's [Characteristic-of-noun-phrase] and [characteristic-of-a-homonym-of-word-1]? A [word-1] [word-2]," filled in by schemas enforcing a noun phrase punchline with appropriate, characteristic descriptions for the setup question. The generation process continuously selects words for every template slot that are consistent with all relations defined in the schema until all slots are validly filled. When JAPE-generated jokes were evaluated by children on a Likert scale from 1 ('not funny at all') to 5 ('very funny'), the most comprehensible ones were found to match the funniness evaluations of published human-generated jokes (Binsted et al., 1997).

Similar lexicon-based approaches have been successfully employed for other types of jokes, such as humorous acronyms generation in the HAHAcronym project (Stock & Strapparava, 2005). Their system created humorous reinterpretations of any given acronym, as well as proposing new fitting acronyms given some keywords. For example, the system changed the expansion of 'FBI' from 'Federal Bureau of Investigation' to 'Fantastic Bureau of Intimidation' (Stock & Strapparava, 2005, 2006).

Initial approaches usually could not discriminate generated jokes based on their perceived funniness. To address this, weights or probabilities could be assigned to particular generation rules to rank the generated jokes heuristically, for example, by using n-gram frequencies or word similarity metrics. A humorous analogy

generator (Petrović & Matthews, 2013) used such weighted generation rules for producing and automatically ranking jokes such as:

I like my coffee like I like my war: cold.

I like my relationships like I like my source: open.

The researchers encoded analogy joke assumptions by requiring the schema to prefer dissimilar nouns for the setup and uncommon, suitable, ambiguous adjectives as a punchline. The joke-generation process fills in the first template slot with values from a human-created analogy joke data set and searches dissimilar nouns and related adjectives for the other two slots. These generated analogies were found to be humorous by humans 16% of the time, compared to 33% for human-created ones. GAG (Generalized Analogy Generator) extended this model by splitting the generation and detection phases, enabling machine learning on rated analogies in a generate-and-test way (Winters et al., 2018). It only outputs randomly generated analogies if the detection model judges the joke to be sufficiently humorous. The jokes generated by this system were evaluated to have a comparable frequency of humorousness as the original model, using fewer assumptions about what constitutes a funnier analogy joke.

Replacing single words in nonhumorous texts is also an effective humor generation method and humor theory testbed. Such systems helped show that taboo words and changing words at the end tend to produce more humorous substitutions (Valitutti et al., 2016). Others showed that explicitly modeling surprise in neural networks for inserting opposing and overlapping context words helps improve pun generation (He et al., 2019).

The advent of large pretrained transformer-based language models such as GPT-2, GPT-3, and BERT enabled revolutionary breakthroughs for most natural language-processing tasks, including computational humor, thanks to their increased linguistic abilities and world knowledge (Brown et al., 2020; Devlin et al., 2019; Radford et al., 2019). One advantage to relying on pretrained language models like GPT-3 for humorous purposes is its ability to mimic text by picking up patterns and repeating these in surprising ways. This is a convenient property for writing comedy sketches, where heightened repetition is a frequently used comedic tool (Besser et al., 2013). For example, GPT-3 generated a comedy sketch where one character keeps describing increasingly weirder images users would submit if one were to host an ‘Ask Me Anything’ on Reddit (Sabeti, 2020). While such examples often involve cherry-picked results, they illustrate the potential of GPT-3 as a powerful brainstorming tool in co-creative collaborations.

The current pretrained language models are more limited in dealing with self-contained jokes and puns. One commonly cited issue for wordplay is that these models encode texts using tokens, which use substrings to map the input to number sequences. The models thus lose information about the exact letters used in the input, making it hard for these models to detect or generate novel wordplay (Branwen, 2020). When fine-tuning GPT-2 on a data set of jokes, it generated mostly absurdist jokes (Frolovs, 2019), such as

What did the chicken say after he got hit by a bus?

"I'm gonna be fine!" (Frolovs, 2019)

The generated jokes are reminiscent of the ones children make, who understand the format of a joke but do not yet understand the formation of a punchline. Similarly, GPT-3 seems to correctly pick up the pattern of given jokes but is usually unable to make the generated text truly humorous (Branwen, 2020). However, for Tom Swifty jokes (jokes in the shape of "Pass me the shellfish," said Tom crabbily.) GPT-3 has been able to come up with somewhat humorous jokes, such as

"I'm having a layover in France, please let me know if you still want me to come over!" Tom said in passing.[4]

Recently several transformer-based architectures were used to generate satirical headlines (Horvitz et al., 2020). Researchers trained multiple document-level BERT models for summarization on a mapping of true headlines, leading paragraphs, and Wikipedia context onto satirical headlines. The outputs of their best BERT-based model was 9.4% of the times perceived as funny (while real *Onion* headlines were 38.4%) and a GPT-2 model fine-tuned on just satirical headlines 6.9%. One of their generated headlines got published after being found more humorous than most of the human-created submissions: "U.S. Asks Pugs If They Can Do Anything" (Horvitz et al., 2020).

Two main issues with humor-generation systems are the lack of annotated data sets and the lack of formal evaluation methods (Hossain et al., 2019). While there are rated edited satirical headline data sets (Hossain et al., 2019; Hossain, Krumm, Sajed, & Kautz, 2020; West & Horvitz, 2019) and rated data sets for other specific joke types (Winters et al., 2018), most large joke data sets[5] lack quality metric values for their the jokes (Mihalcea & Strapparava, 2005). The lack of formal evaluation of the output jokes manifests itself both when evaluating the quality externally, as well as for the system to understand its own humor. There is no standardized methodology or generalizable definition of performance for empirically evaluating and comparing the quality of jokes produced by humor systems (Valitutti et al., 2016). Similarly, most humor systems have no mechanism for trying to understand or check the validity of the humor they produce (He et al., 2019). Learning to automatically detect the humor quality thus potentially can improve humor generation.

Detecting Humor

A computational sense of humor would not only require the ability to generate humor but also to detect when others are making jokes. While humans generally find writing jokes harder than recognizing them (as detection comes naturally), one could argue that automatic generation is easier than automatic detection because the former can focus on narrow subtypes, while a full (or at least larger) coverage of the types of humor would be expected from the latter. One of the first humor detection models used handcrafted humor features with simple models such as naive Bayes and support vector machines to separate one-liners from other types of texts (Mihalcea & Strapparava, 2005). It impressively achieved 97% accuracy for distinguishing one-liners from

news, and 79% accuracy for one-liners from a standard English corpus. While these appear to be incredible results, the question arises whether the model is truly detecting humor or confounding the writing style and lexicon of jokes with their actual humorous qualities (West & Horvitz, 2019; Winters & Delobelle, 2020). Given that the negative examples do not come from the same distribution or domain, these techniques likely learned that certain words occur more often in jokes than other words (e.g., 'bar' or 'mother-in-law'), and thus not capture the nuances of what makes text humorous.

The UR-FUNNY data set includes other modalities than just text (namely, audio and video) to help systems learn about the context through prosodic cues and gestures (Hasan et al., 2019). The researchers constructed the data set using the laughter markers from TED talks as cues to determine which parts lead up to a punchline, and which do not. Using the same source for positive and negative examples helps prevent mistaking style for humor. By training multimodal neural networks on combinations of the modalities, they showed that adding audio and visual cues help improve humor detection accuracy.

While the jokes generated by large-scale transformer models are usually absurd, BERT-like models vastly outperform other approaches in humor detection (Annamoradnejad, 2020; Winters & Delobelle, 2020). This was shown by using a data set containing jokes and generated non-jokes using the same words and structures as real jokes (e.g., "What's the name of Santa's wife? Kitchen table"). Using this data set, typical neural methods like long short-term memory (LSTM) and convolutional neural network (CNN) models are completely unable to distinguish these texts, whereas a RoBERTa-based model achieved almost 90% accuracy on differentiating between the two types (Winters & Delobelle, 2020).

Several efforts have been made for learning the finer nuance in humor detection with the release of multiple parallel rated corpora of similar satirical and real headlines (Hossain et al., 2019; Hossain, Krumm, Sajed, & Kautz, 2020; West & Horvitz, 2019), and organizing competitions around such data sets (Castro et al., 2018; Chiruzzo et al., 2019; Hossain, Krumm, Gamon, & Kautz, 2020). For one such data set, humans edited a single word of a headline to make it funnier (Hossain et al., 2019), while the other started from satire (West & Horvitz, 2019). For the edited headlines, each substitution was accompanied by the average funniness evaluation humans rated this edited headline on a fixed scale. These small edits allow computational models to learn more precisely about the kind of words that make texts more or less funny and even compare the humor level of edits performed on the same headline. During the largest computational humor competition, models were tasked to predict the average rating of each edited headline. As one might expect, all highest ranking teams used large pretrained language models like GPT-2 and BERT to predict the funniness rating (Hossain et al., 2020).

One mostly unaddressed issue in the field of computational humor (both for generation and detection) is how it is mostly centered on English jokes. While there are some humor systems in other languages—such as in Japanese (Terai et al., 2020), Chinese (Chen & Soo, 2018), Spanish (Castro et al., 2016; Castro et al., 2018) and Dutch (Winters, 2019; Winters & Delobelle, 2020)—few researchers evaluated their models on languages

other than English. As humor is hard to translate (Chiaro, 2017; Vandaele, 2010), merely translating the data set (being common practice for natural language processing tasks like textual entailment) will not suffice to create computational humor systems for other languages.

Conclusion

While having a sense of humor is still an elusive trait for machines, many researchers have succeeded in building machines that can generate and detect particular types of humor. With the advent of annotated humor data sets and improved language models, we are also getting closer to the prerequisites of better humor detection and more diverse humor generation. Since recent data sets are also getting increasingly more fine-grained and multimodal, the models will improve inferring the influence of every word and audiovisual cue on the quality of humor.

However, given that a complete sense of humor requires the functional equivalents of most elements of human thought and cognition, building a model that can detect and generate all types of humor remains an AI-complete problem. So, while automated humor models are improving and requiring less handcrafted rules, achieving a true and full computational sense of humor is still far off, meaning that for the foreseeable future, humans will likely still have the last laugh.

Acknowledgments

The author is grateful to the anonymous reviewers and column editor for their insightful comments and suggestions that considerably helped improve the paper, and would also like to thank Benedikte Wallace, Pieter Delobelle and Kory Mathewson for their helpful comments.

Disclosure Statement

Thomas Winters is supported by the Research Foundation-Flanders (FWO-Vlaanderen, 11C7720N).

References

Annamoradnejad, I. (2020). ColBERT: Using BERT sentence embedding for humor detection. *arXiv*. https://doi.org/10.48550/arXiv.2004.12765

Apter, M. (1982). *The experience of motivation: The theory of psychological reversals*. Academic Press. https://doi.org/10.2307/1574929

Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis*. Mouton de Gruyter. https://doi.org/10.1515/9783110887969

Besser, M., Roberts, I., & Walsh, M. (2013). *The Upright Citizens Brigade comedy improvisation manual*. The Comedy Council of Nicea LLC.

Binsted, K. (1995). Using humour to make natural language interfaces more friendly. *Proceedings of the AI, ALife and Entertainment Workshop, International Joint Conference on Artificial Intelligence*. https://www2.hawaii.edu/~binsted/papers/BinstedIJCAI1995.pdf

Binsted, K., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., & O'Mara, D. (2006). Computational humor. *IEEE Intelligent Systems*, *21*(2), 59–69. https://doi.org/10.1007/3-540-47987-2_2

Binsted, K., Pain, H., & Ritchie, G. D. (1997). Children's evaluation of computer-generated punning riddles. *Pragmatics & Cognition*, *5*(2), 305–354. https://doi.org/10.1075/pc.5.2.06bin

Binsted, K., & Ritchie, G. (1994). An implemented model of punning riddles. In *Proceedings of the Twelfth National Conference on Artificial Intelligence/Sixth Conference on Innovative Applications of Artificial Intelligence (AAAI-94)* (pp. 633–638). https://www.aaai.org/Papers/AAAI/1994/AAAI94-096.pdf

Branwen, G. (2020). GPT-3 creative fiction. https://www.gwern.net/GPT-3

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

Caron, J. E. (2002). From ethology to aesthetics: Evolution as a theoretical paradigm for research on laughter, humor, and other comic phenomena. *Humor*, *15*(3), 245–281. https://doi.org/10.1515/humr.2002.015

Castro, S., Chiruzzo, L., & Rosá, A. (2018). Overview of the HAHA task: Humor analysis based on human annotation at IberEval 2018. In P. Rosso, J. Gonzalo, R. Martínez, S. Montalvo, & J. Carrillo-de-Albornoz (Eds.), *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)* (pp. 187–194).

Castro, S., Cubero, M., Garat, D., & Moncecchi, G. (2016). Is this a joke? Detecting humor in Spanish tweets. In M. M. Gómez, H. J. Escalante, A. Segura, & J. de Dios Murillo (Eds.), *Lecture notes in computer science: Vol. 10022. Ibero-American conference on artificial intelligence* (pp. 139–150). https://doi.org/10.1007/978-3-319-47955-2_12

Chan, Y.-C., Chou, T.-L., Chen, H.-C., & Liang, K.-C. (2012). Segregating the comprehension and elaboration processing of verbal jokes: An fMRI study. *Neuroimage*, *61*(4), 899–906.

https://doi.org/10.1016/j.neuroimage.2012.03.052

Chen, P.-Y., & Soo, V.-W. (2018). Humor recognition using deep learning. In M. Walker, H. Ji, & S. Amanda (Eds.), *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)* (pp. 113–117). Association for Computational Linguistics. https://doi.org/10.18653/v1/n18-2018

Chiaro, D. (2017). Humor and translation (S. Attardo, Ed.). In *The Routledge Handbook of Language and Humor* (pp. 414–429). https://doi.org/10.4324/9781315731162-29

Chiruzzo, L., Castro, S., Etcheverry, M., Garat, D., Prada, J. J., & Rosá, A. (2019). Overview of HAHA at IberLEF 2019: Humor analysis based on human annotation. In *Proceedings of the Iberian Languages Evaluation Forum co-located with 35th Conference of the Spanish Society for Natural Language Processing* (pp. 132–144).

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., & Sampath, D. (2010). The YouTube video recommendation system. In X. Amatriain (Ed.), *Proceedings of the Fourth ACM Conference on Recommender Systems* (pp. 293–296). Association for Computing Machinery. https://doi.org/10.1145/1864708.1864770

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423

Frolovs, M. (2019, December 17). Teaching GPT-2 transformer a sense of humor. *Towards Data Science*. https://towardsdatascience.com/teaching-gpt-2-a-sense-of-humor-fine-tuning-large-transformer-models-on-a-single-gpu-in-pytorch-59e8cec40912

Gangal, V., Jhamtani, H., Neubig, G., Hovy, E., & Nyberg, E. (2017). Charmanteau: Character embedding models for portmanteau creation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2917–2922). https://doi.org/10.18653/v1/D17-1315

Gatti, L., Özbal, G., Guerini, M., Stock, O., & Strapparava, C. (2015). Slogans are not forever: Adapting linguistic expressions to the news. In *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 2452–2458).

Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology*, *80*(4), 395–430. https://doi.org/10.1086/498281

Gray, A. W., Parkinson, B., & Dunbar, R. I. (2015). Laughter's influence on the intimacy of self-disclosure. *Human Nature, 26*(1), 28–43. https://doi.org/10.1007/s12110-015-9225-8

Hasan, M. K., Rahman, W., Bagher Zadeh, A., Zhong, J., Tanveer, M. I., Morency, L.-P., & Hoque, M. E. (2019). UR-FUNNY: A multimodal language dataset for understanding humor. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (pp. 2046–2056). https://doi.org/10.18653/v1/D19-1211

He, H., Peng, N., & Liang, P. (2019). Pun generation with surprise. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 1734–1744). https://doi.org/10.18653/v1/N19-1172

Hempelmann, C. F. (2008). Computational humor: Beyond the pun? *The primer of humor research* (pp. 333–360). Mouton de Gruyter. https://doi.org/10.1515/9783110198492.333

Horvitz, Z., Do, N., & Littman, M. L. (2020). Context-driven satirical news generation. In *Proceedings of the Second Workshop on Figurative Language Processing* (pp. 40–50). https://doi.org/10.18653/v1/2020.figlang-1.5

Hossain, N., Krumm, J., & Gamon, M. (2019). "President Vows to Cut ~~Taxes~~ Hair": Dataset and analysis of creative text editing for humorous headlines. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 133–142). https://doi.org/10.18653/v1/N19-1012

Hossain, N., Krumm, J., Gamon, M., & Kautz, H. (2020). SemEval-2020 task 7: Assessing humor in edited news headlines. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 746–758). http://doi.org/10.18653/v1/2020.semeval-1.98

Hossain, N., Krumm, J., Sajed, T., & Kautz, H. (2020). Stimulating creativity with FunLines: A case study of humor generation in headlines. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 256–262). https://doi.org/10.18653/v1/2020.acl-demos.28

Hurley, M. M., Dennett, D. C., Adams Jr, R. B., & Adams, R. B. (2011). *Inside jokes: Using humor to reverse-engineer the mind*. MIT Press. https://doi.org/https://doi.org/10.7551/mitpress/9027.001.0001

Kleinberg, S. (2018, January). *5 ways voice assistance is shaping consumer behavior*. Think with Google. https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/voice-assistance-consumer-experience

Lopatovska, I., Korshakova, E., & Kubert, T. (2020). Assessing user reactions to intelligent personal assistants' humorous responses. In *Proceedings of the Association for Information Science and Technology*, *57*(1), Article e256. https://doi.org/10.1002/pra2.256

Mihalcea, R., & Strapparava, C. (2005). Making computers laugh: Investigations in automatic humor recognition. In R. J. Mooney (Ed.), *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 531–538). Association for Computational Linguistics. https://doi.org/10.3115/1220575.1220642

Morreall, J. (1986). *The philosophy of laughter and humor*. SUNY Press.

Petrovi´c, S., & Matthews, D. (2013). Unsupervised joke generation from big data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 228–232). https://www.aclweb.org/anthology/P13-2041

Preuschoft, S., & van Hooff, J. A. (1997). *The social function of "smile" and "laughter": Variations across primate species and societies.* Lawrence Erlbaum Associates.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*.

Raskin, V. (1984). *Semantic mechanisms of humor*. Springer Netherlands. https://doi.org/10.1007/978-94-009-6472-3_4

Ritchie, G. (1999). Developing the incongruity-resolution theory. In *Proceedings of AISB Symposium on Creative Language: Stories and Humour* (pp. 78–85).

Ritchie, G., Manurung, R., Pain, H., Waller, A., Black, R., & O'Mara, D. (2007). A practical application of computational humour. In A. Cardoso & G. A. Wiggins (Eds.), *Proceedings of the 4th international joint conference on computational creativity* (pp. 91–98). Goldsmiths, University of London.

Sabeti, A. (2020). *Why GPT-3 is good for comedy, or: Don't ever do an AMA on Reddit*. https://arr.am/2020/07/22/why-gpt-3-is-good-for-comedy-or-reddit-eats-larry-page-alive/

Shultz, T. R. (1974). Development of the appreciation of riddles. *Child Development*, *45*(1), 100–105. https://doi.org/10.2307/1127755

Stock, O., & Strapparava, C. (2005). HAHAcronym: A computational humor system. In *Proceedings of the ACL 2005 on Interactive Poster and Demonstration Sessions* (pp. 113–116). https://doi.org/10.3115/1225753.1225782

Stock, O., & Strapparava, C. (2006). Laughing with HAHAcronym, a computational humor system. In *Proceedings of the 21st National Conference on Artificial Intelligence—Volume 2* (pp. 1675–1678). AAAI Press.

Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In J. H. Goldstein & P. E. McGhee (Eds.), *The psychology of humor* (pp. 81–100). Academic Press. https://doi.org/10.1016/B978-0-12-288950-9.50010-9

Terai, A., Yamashita, K., & Komagamine, S. (2020). Computer humor and human humor: Construction of Japanese "nazokake" riddle generation systems. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *24*(2), 199–205. https://doi.org/10.20965/jaciii.2020.p0199

Valitutti, A., Doucet, A., Toivanen, J. M., & Toivonen, H. (2016). Computational generation and dissection of lexical replacement humor. *Natural Language Engineering*, *22*(5), 727–749. https://doi.org/10.1017/s1351324915000145

Vandaele, J. (2010). Humor in translation. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (pp. 147–152). John Benjamins Publishing Company. https://doi.org/10.1075/hts.1.hum1

West, R., & Horvitz, E. (2019). Reverse-engineering satire, or "paper on computational humor accepted despite making serious advances." *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 7265–7272. https://doi.org/10.1609/aaai.v33i01.33017265

Winters, T. (2019). Generating Dutch punning riddles about current affairs. *29th Meeting of Computational Linguistics in the Netherlands (CLIN 2019): Book of Abstracts*.

Winters, T., & Delobelle, P. (2020). Dutch humor detection by generating negative examples. In L. Cao, W. Kosters, & J. Lijffijt (Eds.), *Proceedings of the 32st Benelux Conference on Artificial Intelligence (BNAIC 2020) and the 29th Belgian Dutch Conference on Machine Learning (Benelearn 2020)*. Universiteit Leiden.

Winters, T., Nys, V., & De Schreye, D. (2018). Automatic joke generation: Learning humor from examples. In N. Streitz, & S. Konomi (Eds.), *Lecture notes in computer science: Vol. 10922. Distributed, ambient and pervasive interactions: Technologies and contexts* (pp. 360–377). https://doi.org/10.1007/978-3-319-91131-1_28

Wyer, R., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, *99(4)*, 663–688. https://doi.org/10.1037/0033-295x.99.4.663

---

article.

Footnotes

1. https://www.thegag.club/ ↩
2. Formerly https://pitch.live, but now defunct ↩
3. https://www.grammarly.com/ ↩
4. From https://www.gwern.net/GPT-3#tom-swifties ↩
5. Most famously the *Short Jokes* data set containing 200K+ jokes: https://www.kaggle.com/abhinavmoudgil95/short-jokes ↩

References

- Annamoradnejad, I. (2020). ColBERT: Using *bert* sentence embedding for humor detection. *ArXiv*. https://arxiv.org/abs/2004.12765 ↩
- Apter, M. (1982). *The experience of motivation: The theory of psychological reversals*. Academic Press. https://doi.org/10.2307/1574929 ↩
- Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis*. Mouton de Gruyter. https://doi.org/10.1515/9783110887969 ↩
- Besser, M., Roberts, I., & Walsh, M. (2013). *The Upright Citizens Brigade comedy improvisation manual*. The Comedy Council of Nicea LLC. ↩
- Binsted, K. (1995). Using humour to make natural language interfaces more friendly. *Proceedings of the Ai, Alife and Entertainment Workshop, Intern. Joint Conf. On Artificial Intelligence*. https://www2.hawaii.edu/~binsted/papers/BinstedIJCAI1995.pdf ↩
- Binsted, K., & Ritchie, G. (1994). An implemented model of punning riddles. *Proceedings of the Twelfth National Conference on Artificial Intelligence/Sixth Conference on Innovative Applications of Artificial Intelligence (AAAI-94)*, 633–638. https://www.aaai.org/Papers/AAAI/1994/AAAI94-096.pdf ↩
- Binsted, K., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., & O'Mara, D. (2006). Computational humor. *IEEE Intelligent Systems*, *21*(2), 59–69. https://doi.org/10.1007/3-540-47987-2_2 ↩
- Binsted, K., Pain, H., & Ritchie, G. D. (1997). Children's evaluation of computer-generated punning riddles. *Pragmatics & Cognition*, *5*(2), 305–354. https://doi.org/10.1075/pc.5.2.06bin ↩
- Branwen, G. (2020). *GPT-3 creative fiction*. https://www.gwern.net/GPT-3 ↩
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & others. (2020). Language models are few-shot learners. *ArXiv*. https://arxiv.org/abs/2005.14165 ↩

- Caron, J. E. (2002). From ethology to aesthetics: Evolution as a theoretical paradigm for research on laughter, humor, and other comic phenomena. *Humor*, *15*(3), 245–281. https://doi.org/10.1515/humr.2002.015 ↩
- Castro, S., Chiruzzo, L., & Rosá, A. (2018). Overview of the HAHA task: Humor analysis based on human annotation at IberEval 2018. In P. Rosso, J. Gonzalo, R. Martínez, S. Montalvo, & J. Carrillo-de Albornoz (Eds.), *Proceedings of the third workshop on evaluation of human language technologies for iberian languages (ibereval 2018) co-located with 34th conference of the spanish society for natural language processing (sepln 2018)* (pp. 187–194). ↩
- Castro, S., Cubero, M., Garat, D., & Moncecchi, G. (2016). Is this a joke? Detecting humor in spanish tweets. *Ibero-American Conference on Artificial Intelligence*, 139–150. https://doi.org/10.1007/978-3-319-47955-2_12 ↩
- Chan, Y.-C., Chou, T.-L., Chen, H.-C., & Liang, K.-C. (2012). Segregating the comprehension and elaboration processing of verbal jokes: An fMRI study. *Neuroimage*, *61*(4), 899–906. https://doi.org/10.1016/j.neuroimage.2012.03.052 ↩
- Chen, P.-Y., & Soo, V.-W. (2018). Humor recognition using deep learning. In M. Walker, H. Ji, & S. Amanda (Eds.), *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)* (pp. 113–117). Association for Computational Linguistics. https://doi.org/10.18653/v1/n18-2018 ↩
- Chiaro, D. (2017). Humor and translation. *The Routledge Handbook of Language and Humor*, 414–429. https://doi.org/10.4324/9781315731162-29 ↩
- Chiruzzo, L., Castro, S., Etcheverry, M., Garat, D., Prada, J. J., & Rosá, A. (2019). Overview of HAHA at IberLEF 2019: Humor analysis based on human annotation. *Proceedings of the Iberian Languages Evaluation Forum Co-Located with 35th Conference of the Spanish Society for Natural Language Processing*, 132–144. ↩
- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., & others. (2010). The YouTube video recommendation system. In X. Amatriain (Ed.), *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293–296). Association for Computing Machinery. ↩
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. https://doi.org/10.18653/v1/N19-1423 ↩
- Frolovs, M. (2019). Teaching GPT-2 transformer a sense of humor. *Towards Data Science*. https://towardsdatascience.com/teaching-gpt-2-a-sense-of-humor-fine-tuning-large-transformer-models-on-a-single-gpu-in-pytorch-59e8cec40912 ↩
- Gangal, V., Jhamtani, H., Neubig, G., Hovy, E., & Nyberg, E. (2017). Charmanteau: Character embedding models for portmanteau creation. *Proceedings of the 2017 Conference on Empirical Methods in Natural*

Language Processing, 2917–2922. https://doi.org/10.18653/v1/D17-1315↩

- Gatti, L., Özbal, G., Guerini, M., Stock, O., & Strapparava, C. (2015). Slogans are not forever: Adapting linguistic expressions to the news. *Proceedings of the 24th International Conference on Artificial Intelligence*, 2452–2458. ↩
- Gervais, M., & Wilson, D. S. (2005). The evolution and functions of laughter and humor: A synthetic approach. *The Quarterly Review of Biology*, *80*(4), 395–430. https://doi.org/10.1086/498281 ↩
- Gray, A. W., Parkinson, B., & Dunbar, R. I. (2015). Laughter's influence on the intimacy of self-disclosure. *Human Nature*, *26*(1), 28–43. https://doi.org/10.1007/s12110-015-9225-8 ↩
- Hasan, M. K., Rahman, W., Bagher Zadeh, A., Zhong, J., Tanveer, M. I., Morency, L.-P., & Hoque, M. E. (2019). UR-FUNNY: A multimodal language dataset for understanding humor. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (Emnlp-Ijcnlp)*, 2046–2056. https://doi.org/10.18653/v1/D19-1211 ↩
- He, H., Peng, N., & Liang, P. (2019). Pun generation with surprise. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 1734–1744. https://doi.org/10.18653/v1/N19-1172 ↩
- Hempelmann, C. F. (2008). Computational humor: Beyond the pun? In *The primer of humor research* (pp. 333–360). Mouton de Gruyter. https://doi.org/10.1515/9783110198492.333 ↩
- Horvitz, Z., Do, N., & Littman, M. L. (2020). Context-driven satirical news generation. *Proceedings of the Second Workshop on Figurative Language Processing*, 40–50. https://doi.org/10.18653/v1/2020.figlang-1.5 ↩
- Hossain, N., Krumm, J., & Gamon, M. (2019). "President vows to cut \<Taxes\> hair": Dataset and analysis of creative text editing for humorous headlines. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 133–142. https://doi.org/10.18653/v1/N19-1012 ↩
- Hossain, N., Krumm, J., & Gamon, M. (2019). "President Vows to Cut Taxes Hair": Dataset and analysis of creative text editing for humorous headlines. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 133–142. https://doi.org/10.18653/v1/N19-1012 ↩
- Hossain, N., Krumm, J., Gamon, M., & Kautz, H. (2020). SemEval-2020 task 7: Assessing humor in edited news headlines. *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, 746–758. ↩
- Hossain, N., Krumm, J., Sajed, T., & Kautz, H. (2020). Stimulating creativity with FunLines: A case study of humor generation in headlines. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 256–262. https://doi.org/10.18653/v1/2020.acl-demos.28 ↩
- Hurley, M. M., Dennett, D. C., Adams Jr, R. B., & Adams, R. B. (2011). *Inside jokes: Using humor to reverse-engineer the mind*. MIT press. https://doi.org/https://doi.org/10.7551/mitpress/9027.001.0001 ↩

- Kleinberg, S. (2018). 5 ways voice assistance is shaping consumer behavior. In *Think with Google*. Google. https://www.thinkwithgoogle.com/future-of-marketing/emerging-technology/voice-assistance-consumer-experience/ ↩
- Lopatovska, I., Korshakova, E., & Kubert, T. (2020). Assessing user reactions to intelligent personal assistants' humorous responses. *Proceedings of the Association for Information Science and Technology*, *57*(1), Article e256. https://doi.org/10.1002/pra2.256 ↩
- Mihalcea, R., & Strapparava, C. (2005). Making computers laugh: Investigations in automatic humor recognition. In R. Joseph Mooney (Ed.), *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 531–538). Association for Computational Linguistics; Association for Computational Linguistics. https://doi.org/10.3115/1220575.1220642 ↩
- Morreall, J. (1986). *The philosophy of laughter and humor*. SUNY Press. ↩
- Petrović, S., & Matthews, D. (2013). Unsupervised joke generation from big data. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 228–232. https://www.aclweb.org/anthology/P13-2041 ↩
- Preuschoft, S., & Hooff, van J. A. R. A. M. (1997). *The social function of "smile" and "laughter": Variations across primate species and societies.* (pp. 171–190). Lawrence Erlbaum Associates. ↩
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*. ↩
- Raskin, V. (1984). *Semantic mechanisms of humor*. Springer Netherlands. https://doi.org/10.1007/978-94-009-6472-3_4 ↩
- Ritchie, G. (1999). Developing the incongruity-resolution theory. *Proceedings of AISB Symposium on Creative Language: Stories and Humour*, 78–85. ↩
- Ritchie, G., Manurung, R., Pain, H., Waller, A., Black, R., & O'Mara, D. (2007). A practical application of computational humour. In A. Cardoso & G. A. Wiggins (Eds.), *Proceedings of the 4th international joint conference on computational creativity* (pp. 91–98). Goldsmiths, University of London. ↩
- Sabeti, A. (2020). *Why GPT-3 is good for comedy, or: Don't ever do an AMA on Reddit*. https://arr.am/2020/07/22/why-gpt-3-is-good-for-comedy-or-reddit-eats-larry-page-alive/ ↩
- Shultz, T. R. (1974). Development of the appreciation of riddles. *Child Development*, *45*(1), 100–105. https://doi.org/10.2307/1127755 ↩
- Stock, O., & Strapparava, C. (2005). HAHAcronym: A computational humor system. *Proceedings of the Acl 2005 on Interactive Poster and Demonstration Sessions*, 113–116. https://doi.org/10.3115/1225753.1225782 ↩
- Stock, O., & Strapparava, C. (2006). Laughing with hahacronym, a computational humor system. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2* (pp. 1675–1678). AAAI Press. ↩
- Suls, J. M. (1972). A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In J. H. Goldstein & P. E. McGhee (Eds.), *The psychology of humor* (pp. 81–100). Academic Press.

https://doi.org/https://doi.org/10.1016/B978-0-12-288950-9.50010-9↩

- Terai, A., Yamashita, K., & Komagamine, S. (2020). Computer humor and human humor: Construction of japanese “nazokake” riddle generation systems. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *24*(2), 199–205. https://doi.org/10.20965/jaciii.2020.p0199 ↩
- Valitutti, A., Doucet, A., Toivanen, J. M., & Toivonen, H. (2016). Computational generation and dissection of lexical replacement humor. *Natural Language Engineering*, *1*(1), 1–24. https://doi.org/10.1017/s1351324915000145 ↩
- Vandaele, J. (2010). Humor in translation. In Y. Gambier & van L. Doorslaer (Eds.), *Handbook of translation studies* (Vol. 1, pp. 147–152). John Benjamins Publishing Company. https://doi.org/10.1075/hts.1.hum1 ↩
- West, R., & Horvitz, E. (2019). Reverse-engineering satire, or “paper on computational humor accepted despite making serious advances.” *Proceedings of the Aaai Conference on Artificial Intelligence*, *33*, 7265–7272. https://doi.org/10.1609/aaai.v33i01.33017265 ↩
- Winters, T. (2019). Generating Dutch punning riddles about current affairs. *29th Meeting of Computational Linguistics in the Netherlands (CLIN 2019): Book of Abstracts*. ↩
- Winters, T., & Delobelle, P. (2020). Dutch humor detection by generating negative examples. In L. Cao, W. Kosters, & J. Lijffijt (Eds.), *Proceedings of the 32st Benelux Conference on Artificial Intelligence (BNAIC 2020) and the 29th Belgian Dutch Conference on Machine Learning (Benelearn 2020)*. Universiteit Leiden. ↩
- Winters, T., & Delobelle, P. (2020). Dutch humor detection by generating negative examples. In L. Cao, W. Kosters, & J. Lijffijt (Eds.), *Proceedings of the 32st Benelux Conference on Artificial Intelligence (BNAIC 2020) and the 29th Belgian Dutch Conference on Machine Learning (Benelearn 2020)*. Universiteit Leiden. ↩
- Winters, T., & Delobelle, P. (2020). Dutch humor detection by generating negative examples. In L. Cao, W. Kosters, & J. Lijffijt (Eds.), *Proceedings of the 32st Benelux Conference on Artificial Intelligence (BNAIC 2020) and the 29th Belgian Dutch Conference on Machine Learning (Benelearn 2020)*. Universiteit Leiden. ↩
- Winters, T., Nys, V., & De Schreye, D. (2018). Automatic joke generation: Learning humor from examples. *Distributed, Ambient and Pervasive Interactions: Technologies and Contexts*, *10922 LNCS*, 360–377. https://doi.org/10.1007/978-3-319-91131-1_28 ↩
- Wyer, R., & Collins, J. E. (1992). A theory of humor elicitation. *Psychological Review*, *99(4)*, 663–688. https://doi.org/10.1037/0033-295x.99.4.663 ↩