How LLMs Reinforce Political Misinformation: Insights from the Analysis of False Presuppositions

Judith Sieker, Clara Lachenmaier and Sina Zarrieß
Computational Linguistics, Department of Linguistics
Bielefeld University, Germany
{j.sieker, clara.lachenmaier, sina.zarriess}@uni-bielefeld.de

1 Introduction

This paper explores the potential of large language models (LLMs) to amplify misinformation by analyzing their responses to user queries through the lens of linguistic presupposition analysis.

In recent years, the rise of populist politicians and the spread of fake news (by these populists) have intensified. Fake news is said to have democracy-destroying effects by polarizing voters, dominating the public debate, and undermining traditional media (Curini and Pizzimenti, 2020). As LLMs play an increasing role in public discourse, it is crucial to understand their potential in shaping political opinions. One relevant factor is the sycophancy effect, which shows that LLMs often tend to adjust their responses to align with users' views (Perez et al., 2023), raising concerns about how these models might reinforce misinformation. Therefore, it is crucial to investigate how LLMs respond when prompts reflect that users have fallen for misinformation due to their political biases.

Our study employs a systematic methodology based on linguistic presupposition analysis to examine this question for three temporary AI language models. Specifically, our study focuses on false presuppositions – presuppositions where the proposition assumed to be true is instead false (Yablo, 2006) – and examines the conditions under which LLMs are more likely to accept and reinforce them. We conduct two experiments with two different datasets (one newly created), exploring factors like linguistic constructions, embedding context, and scenario probability. Preliminary observations suggest that contemporary LLMs have difficulty recognizing false presuppositions, with their performance varying depending on the conditions.

This paper contributes to the understanding of LLMs in the context of political misinformation by: (1) revealing their vulnerability to biased misinformation and false presuppositions; (2) introducing a new framework for testing false presuppositions; and (3) developing a dataset to study political misinformation and false presuppositions in LLMs.

2 Background

Speakers often rely on presuppositions, i.e., underlying assumptions or shared beliefs that are taken for granted in communication (Stalnaker, 1973). Presupposition triggers, such as possessives or quantifiers, introduce these assumptions and are common in everyday language (Beaver et al., 2024). An intriguing aspect of presuppositions, which our study focuses on, is presupposition failure, where a presupposed fact is actually false (Yablo, 2006). For example, "the prime minister of Germany" presupposes that Germany has a prime minister, which fails because it does not. Such failures can disrupt communication due to missing background information (Xia et al., 2019). However, failure does not always disrupt discourse; hearers may adjust their beliefs to align with the presupposition, a process called accommodation. E.g., if someone says "my dog" to an audience unaware of their pet, listeners may accommodate by assuming the speaker has a dog. Whether accommodation occurs depends on the hearers' willingness to adjust their understanding based on the context (Von Fintel, 2008).

Similarly, when LLMs encounter (false) presuppositions, they must determine whether to accommodate the information. So far, prior research on presuppositions in LLMs has largely focused on classification tasks, revealing that LLMs often identify surface patterns but overlook deeper knowledge (Jeretic et al., 2020; Kabbara and Cheung, 2022; Cong, 2022). Furthermore, studies on presupposition failure, especially in question-answering systems, indicate that these models struggle with false presuppositions (Kim et al., 2021, 2023; Daswani et al., 2024; Yu et al., 2023; Srikanth et al., 2024). These findings raise important questions about how

advanced models like GPT-4 (Achiam et al., 2023) – known for its strong performance in various linguistic tasks (Cai et al., 2023) – as well as other models like LLama (Dubey et al., 2024) and Mistral (Jiang et al., 2023) manage false presuppositions in more complex contexts. Our study addresses this gap by examining these models specifically in political contexts, where the mismanagement of false information could have serious consequences. We conduct two experiments using datasets and conditions that align with linguistic analyses of presuppositions, addressing gaps in prior work that focused on examining false presuppositions in questions.

3 Methodology

We employ two experiments to investigate how LLMs reinforce misinformation by analyzing their tendency to identify and accommodate false presuppositions. For both experiments, we examine OpenAI's GPT-4-o (Achiam et al., 2023), MistralAI's Mistral-7B-v03 (Jiang et al., 2023), and Meta's LLama-3-8B (Dubey et al., 2024), all instruction-tuned models. In each experiment, each model is tested with the same prompt three times. Furthermore, we focus on the four parties, LINKE, SPD, CDU/CSU, and AfD, to represent a balanced distribution on the left-right scale.

Experiment 1: Wahl-O-Mat. In this experiment, we used data from the Wahl-O-Mat,[1] a German tool for comparing political views with party positions. The dataset includes 38 statements from the 2024 European elections, with party responses showing their (dis-)agreement. To test whether LLMs recognize false presuppositions, we incorporated these statements into sentences with six different factive verbs as presupposition triggers. For each statement and party, we created both "true" and "false" presuppositions. For example, if the CDU/CSU agreed on a statement about the Euro, a true presupposition might be: "Did the voters find out that CDU/CSU wants to abolish the Euro?" The false presupposition would be: "Did the voters find out that CDU/CSU does not want to abolish the Euro?" All questions were polar questions to simplify evaluation, and we tested 1,104 prompts.

Experiment 2: Presuppositional factors. In this experiment, we analyze how three factors – trigger type, embedding context (question, negation, modal), and scenario probability – affect LLMs' ability to recognize false presuppositions. Due to the complexity of these factors, we created a custom dataset. In each prompt, a politician from one party was inaccurately reported to be active at another party's conference. We used seven types of presupposition triggers, such as factive verbs and quantifiers, with 23 individual triggers. We examined three embedding contexts (questions, negations, modals) and varied the scenario probability with both "probable" and "improbable" events to happen at a party's conference, drawing on insights from (psycho-)linguistic research on presupposition behavior (Tonhauser et al., 2018; Tonhauser, 2016; Mahler, 2020; Degen and Tonhauser, 2021). In this way, we created 1,104 prompts for testing.

4 Evaluation and First Observations

For evaluation, the authors personally annotate the models' responses, which proved to be more complex than initially anticipated, requiring linguistic as well as political expertise. For example, polar questions often received answers beyond simple yes or no, requiring a thorough reading of the text.

Preliminary observations indicate that contemporary LLMs have difficulty recognizing false presuppositions, with their performance varying based on specific conditions. For instance, the type of presupposition trigger seems to influence whether a false presupposition is accommodated. Additionally, the probability of the scenario seems to impact detection: for example, models seem to be more successful at identifying false presuppositions related to core party issues, such as supportive immigration statements from right-wing parties like the AfD. These initial observations align with our expectations derived from (psycho-)linguistic research. Furthermore, it appears that the individual models exhibit inconsistent behavior, as repeating a prompt three times often yields varying responses.

Our preliminary observations suggest that linguistic presupposition analysis is a valuable tool for assessing LLMs' susceptibility to reinforcing political misinformation. I.e., by examining how LLMs handle (false) presuppositions, especially in politically charged contexts, we can gain insights into how these models might reflect or amplify biases in their outputs. At the workshop, we aim to discuss these preliminary findings in detail, along with the final results and their implications for using LLMs in political contexts.

[1] https://www.bpb.de/themen/wahl-o-mat

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

David I. Beaver, Bart Geurts, and Kristie Denlinger. 2024. Presupposition. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Fall 2024 edition. Metaphysics Research Lab, Stanford University.

Zhenguang Garry Cai, David Haslett, Xufeng Duan, Shuqi Wang, and Martin John Pickering. 2023. Does chatgpt resemble humans in language use?

Yan Cong. 2022. Psycholinguistic diagnosis of language models' commonsense reasoning. In *Proceedings of the First Workshop on Commonsense Representation and Reasoning (CSRR 2022)*, pages 17–22, Dublin, Ireland. Association for Computational Linguistics.

Luigi Curini and Eugenio Pizzimenti. 2020. Searching for a unicorn: Fake news and electoral behaviour. *Democracy and Fake News*, pages 77–91.

Ashwin Daswani, Rohan Sawant, and Najoung Kim. 2024. Syn-qa2: Evaluating false assumptions in long-tail questions with synthetic qa datasets. *Preprint*, arXiv:2403.12145.

Judith Degen and Judith Tonhauser. 2021. Prior Beliefs Modulate Projection. *Open Mind*, 5:59–70.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaoqing Ellen Tan, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aaron Grattafiori, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alex Vaughan, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Franco, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, Danny Wyatt, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Mont-

gomery, Eleonora Presani, Emily Hahn, Emily Wood, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Firat Ozgenel, Francesco Caggioni, Francisco Guzmán, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Govind Thattai, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Igor Molybog, Igor Tufanov, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Karthik Prasad, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kun Huang, Kunal Chawla, Kushal Lakhotia, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Maria Tsimpoukelli, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikolay Pavlovich Laptev, Ning Dong, Ning Zhang, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Rohan Maheswari, Russ Howes, Ruty Rinott, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Kohler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaofang Wang, Xiaojian Wu, Xiaolan Wang, Xide Xia, Xilun Wu, Xinbo Gao, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yuchen Hao, Yundi Qian, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, and Zhiwei Zhao. 2024. The llama 3 herd of models. *Preprint*, arXiv:2407.21783.

Paloma Jeretic, Alex Warstadt, Suvrat Bhooshan, and Adina Williams. 2020. Are natural language inference models IMPPRESsive? Learning IMPlicature and PRESupposition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8690–8705, Online. Association for Computational Linguistics.

Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.

Jad Kabbara and Jackie Chi Kit Cheung. 2022. Investigating the performance of transformer-based NLI models on presuppositional inferences. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 779–785, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Najoung Kim, Phu Mon Htut, Samuel R. Bowman, and Jackson Petty. 2023. $(QA)^2$: Question answering with questionable assumptions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8466–8487, Toronto, Canada. Association for Computational Linguistics.

Najoung Kim, Ellie Pavlick, Burcu Karagol Ayan, and Deepak Ramachandran. 2021. Which linguist invented the lightbulb? presupposition verification for question-answering. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3932–3945, Online. Association for Computational Linguistics.

Taylor Mahler. 2020. The social component of the projection behavior of clausal complement contents. *Proceedings of the Linguistic Society of America*, 5(1):777–791.

Ethan Perez, Sam Ringer, Kamile Lukosiute, Karina Nguyen, Edwin Chen, Scott Heiner, Craig Pettit, Catherine Olsson, Sandipan Kundu, Saurav Kadavath, Andy Jones, Anna Chen, Benjamin Mann, Brian Israel, Bryan Seethor, Cameron McKinnon, Christopher Olah, Da Yan, Daniela Amodei, Dario Amodei, Dawn Drain, Dustin Li, Eli Tran-Johnson, Guro Khundadze, Jackson Kernion, James Landis, Jamie Kerr, Jared Mueller, Jeeyoon Hyun, Joshua Landau, Kamal Ndousse, Landon Goldberg, Liane

Lovitt, Martin Lucas, Michael Sellitto, Miranda Zhang, Neerav Kingsland, Nelson Elhage, Nicholas Joseph, Noemi Mercado, Nova DasSarma, Oliver Rausch, Robin Larson, Sam McCandlish, Scott Johnston, Shauna Kravec, Sheer El Showk, Tamera Lanham, Timothy Telleen-Lawton, Tom Brown, Tom Henighan, Tristan Hume, Yuntao Bai, Zac Hatfield-Dodds, Jack Clark, Samuel R. Bowman, Amanda Askell, Roger Grosse, Danny Hernandez, Deep Ganguli, Evan Hubinger, Nicholas Schiefer, and Jared Kaplan. 2023. Discovering language model behaviors with model-written evaluations. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13387–13434, Toronto, Canada. Association for Computational Linguistics.

Neha Srikanth, Rupak Sarkar, Heran Mane, Elizabeth Aparicio, Quynh Nguyen, Rachel Rudinger, and Jordan Boyd-Graber. 2024. Pregnant questions: The importance of pragmatic awareness in maternal health question answering. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7253–7268, Mexico City, Mexico. Association for Computational Linguistics.

Robert Stalnaker. 1973. Presuppositions. *Journal of Philosophical Logic*, 2(4):447–457.

Judith Tonhauser. 2016. Prosodic cues to presupposition projection. *Semantics and Linguistic Theory*, 26:934.

Judith Tonhauser, David I Beaver, and Judith Degen. 2018. How Projective is Projective Content? Gradience in Projectivity and At-issueness. *Journal of Semantics*, 35(3):495–542.

Kai Von Fintel. 2008. What is presupposition accommodation, again? *Philosophical Perspectives*, 22(1):137–170.

Alice Xia, Roxana M Barbu, Kathleen Van Benthem, Daniel A Di Giovanni, Ida Toivonen, and Raj Singh. 2019. Detecting presupposition failure and accommodation with EEG. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 41(0).

S Yablo. 2006. Non-catastrophic presupposition failure.

Xinyan Yu, Sewon Min, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2023. CREPE: Open-domain question answering with false presuppositions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10457–10480, Toronto, Canada. Association for Computational Linguistics.