Multilingual Relative Clause Attachment Ambiguity Resolution in Large Language Models

So Young Lee[•], Russell Scheinberg[◇], Amber Shore[◇], Ameeta Agrawal[◇]

[•]Miami University, USA
[◇]Portland State University, USA
soyoung.lee@miamioh.edu, rschein2@pdx.edu, ashore@pdx.edu, ameeta@pdx.edu

Abstract

This study examines how large language models (LLMs) resolve relative clause (RC) attachment ambiguities and compares their performance to human sentence processing. Focusing on two linguistic factors, namely the length of RCs and the syntactic position of complex determiner phrases (DPs), we assess whether LLMs can achieve human-like interpretations amid the complexities of language. In this study, we evaluated several LLMs, including Claude, Gemini and Llama, in multiple languages: English, Spanish, French, German, Japanese, and Korean. While these models performed well in Indo-European languages (English, Spanish, French, and German), they encountered difficulties in Asian languages (Japanese and Korean), often defaulting to incorrect English translations. The findings underscore the variability in LLMs' handling of linguistic ambiguities and highlight the need for model improvements, particularly for non-European languages. This research informs future enhancements in LLM design to improve accuracy and human-like processing in diverse linguistic environments.

1 Introduction

The primary objective of Natural Language Processing (NLP) research is to design language models that can interpret and generate language in the same way humans do. This is particularly challenging due to the complexity and nuance of human language, which often includes idiomatic expressions, context-dependent meanings, and subtle variations in tone and intent. Additionally, ambiguity in human language, where words and phrases can have multiple interpretations, further complicates the task of these models.

Ambiguity is critical in human-computer interaction due to its pervasiveness in everyday life. Failure to correctly interpret a user's intentions can cause the user to mistrust the system and discontinue use. For decades, ambiguity, therefore, has been a challenging issue for NLP researchers (Davis and van Schijndel, 2020). Despite some progress in resolving ambiguity problems, it still remains a significant challenge for computational linguists and computer scientists.

Recent advances in large language models (LLMs) have significantly improved their ability to process and generate language (Team et al., 2024; Dubey et al., 2024). However, can these LLMs also handle ambiguity well? In human language, ambiguity appears at various levels, one of which is syntactic ambiguity, which occurs when a sentence can be analyzed as having more than one syntactic structure or parse tree as in (1).

(1) The girl saw the boy with the binoculars.
  a. VP modification: The girl used the binoculars to see the boy.
  b. NP modification: The boy had the binoculars, and the girl saw him.

So far, extensive research has been conducted in NLP to address ambiguity. However, the majority of this research has centered on resolving prepositional phrase (PP) attachment ambiguity only (Yin et al., 2021; Xin et al., 2021). Despite its frequent discussion within the field of psycholinguistics, there has been surprisingly little research specifically on **relative clause (RC) attachment ambiguity**, which also can happen in human-computer interaction as in (2).

(2) Play the cover$_{DP1}$ of the song$_{DP2}$ [that features the famous violinist]$_{RC}$.
  a. DP1 modification: The user wants to hear a cover version of a song that specifically includes the participation of the famous violinist.
  b. DP2 modification: The user is asking to play a cover version of a specific

song known for featuring a famous violinist.

Consequently, it is necessary to expand our scope to comprehensively evaluate how LLMs handle various syntactic attachment ambiguities. In this study, we aim to explore how the most recently developed and widely used LLMs resolve RC attachment ambiguities. The assessment of LLMs' performance on RC attachment ambiguities provides insight into the current advancements in language model development.

Our key contributions are as follows:

- We focus on a well-defined linguistic phenomenon and explore how (four recently introduced) LLMs can be effectively prompted to identify relative clauses (RC) and how they handle specific RC attachment ambiguities, comparing their performance with human experimental data.

- Our study extends the RC attachment ambiguity experiment across multiple languages, including European languages, Japanese, and Korean[1], highlighting the variation in LLM performance across different linguistic contexts.

- We extend the existing dataset to two new languages: Japanese and Korean, which will be made available to support further research.

2 Related Work

2.1 Findings in Psycholinguistics

Consider a sentence in (2) again. When a complex determiner phrase (DP) of the form *DP1 of DP2* is followed by an RC, ambiguity arises.

As shown in Table 1, languages exhibit varying preferences for attaching RCs to one of two potential DPs — either DP1: *the cover*, or DP2: *the song*. This leads to either High Attachment (HA; where the RC modifies the first DP which is non-local) or Low Attachment (LA; where the RC modifies the second DP which is local) interpretations (see Figure 1) (Cuetos and Mitchell, 1988; Carreiras and Clifton Jr, 1993, a.o.).

[1]Code and data available at `https://github.com/PortNLP/Multilingual_RC_Attachment/`.

[2]Both HA and LA preferences were reported in German and Portuguese Hemforth et al. (1996); Augurzky (2006) Japanese and Korean are added into the table in, based on the results in Kamide and Mitchell (1997); Lee (2021)

Low Attachment	**High Attachment**
Arabic	Afrikaans
Basque	Bulgarian
Bulgarian	Serbo-Croatian
Chinese	Dutch
English	French
German	Galician
Norwegian	German
Portuguese	Greek
Romanian	Italian
Swedish	Japanese
	Korean
	Portuguese
	Russian
	Spanish

Table 1: Summary of Language Preferences for Relative Clause Attachment (Grillo and Costa, 2014). Languages that exhibit both low and high attachment preferences are underlined[2].

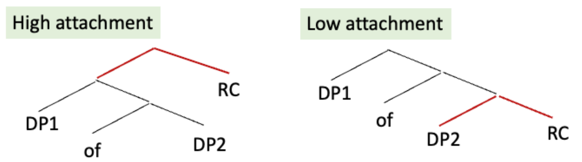

Figure 1: Syntactic structures for the two interpretations

Additionally, within the same language, variations in attachment preferences have been reported, which suggests that factors such as locality, frequency, syntactic position, semantic or pragmatic plausibility, and implicit prosody play significant roles in ambiguity resolution (Gilboy et al., 1995; Acuna-Farina et al., 2009; Fernández, 2005; Fraga et al., 2005, a.o). Among many, the length of constituents such as RCs is a crucial factor. According to Fodor (1998)'s Balanced Sister hypothesis, constituents like RCs, preferentially attach to elements of similar weight or length to maintain prosodic balance. For example, a lengthy RC such as *who frequently attended lavish court gatherings* is more likely to attach to a higher-level constituent, such as *the son of the king*, to preserve prosodic harmony. In contrast, a shorter RC, like *who drank*, tends to attach to a lower-level constituent, such as *the king*, to achieve this balance.

Focus also plays a pivotal role in resolving attachment ambiguities. Schafer (1996) demonstrated that a pitch accent on a noun within a DP influences the attachment of a RC to that noun.

The placement of nouns within a sentence often correlates with their focus; for instance, Carlson et al. (2009) described the 'nuclear-scope'—the typical site for asserted or focused information—as including object positions but not preverbal or initial subject positions, which are typically associated with topical or previously known information. Previous studies show a distinction in RC attachment between object and non-object positions; in object positions, the DP usually receives broad focus, making the first DP the likely attachment site for the RC (Hemforth and Konieczny, 2002).

Hemforth et al. (2015) reports the effects of the length of RCs and the position of complex DPs in four different languages, English, French, German, and Spanish but this effect varies across those languages. As shown in Figure 2, French results showed overall HA preference regardless of DP positions, which differed from English showing LA. In German and Spanish, preferences for HA and LA varied based on specific conditions. Overall, it was observed that long RCs increased the percentage of HA choices, particularly in object positions, suggesting a role for implicit prosodic phrasing. This increase was especially pronounced in German and Spanish. In addition, RCs in object positions demonstrated a greater tendency for HA compared to those in subject/topic positions.

These findings resonate with known patterns in human sentence processing, where prosodic cues and syntactic structures serve as heuristics to resolve ambiguities. The observed language-specific variations in attachment preferences indicate that these heuristics are tailored to the unique structural and prosodic environments of each language.

2.2 Findings in NLP

Although studies on RC attachment ambiguity in NLP have been rare, numerous investigations have used RCs to examine the syntactic structures represented in language models (LMs), using either synthetic or naturalistic data to determine if LMs represent specific linguistic features or biases (Prasad and Linzen, 2024). For instance, Prasad et al. (2019) tested structural priming on pre-Transformer long short-term memory (LSTM) neural networks by adapting these models to different types of RCs and non-RC sentences. They found that models adapted to a specific RC type showed reduced surprisal to sentences of that RC type compared to other RC types, and reduced surprisal to RC sentences in general compared to non-RC sentences. This suggests that LSTM models develop hierarchical syntactic representations. Prior work has also examined LMs (BERT, RoBERTa, and ALBERT) for sentence-level syntactic and semantic understanding (Warstadt and Bowman, 2019; Mosbach et al., 2020). These studies found that while these models perform well in parsing syntactic information, they struggle to predict masked relative pronouns using context and semantic knowledge.

The discussion initially focused on English, but it gradually expanded to explore how the performance of LMs manifests in other languages. Tikhonova et al. (2023) on multilingual BERT (mBERT) examined how well it understands and processes linguistic structures, including RCs, through the natural language inference task. It found that extra data in English improves stability for all other tested languages (French, German, Russian, Swedish).

The most relevant work to our study is Davis and van Schijndel (2020), which explored the linguistic biases of RNN-based language models in resolving RC attachment ambiguity. This research specifically examined how these models handle HA and LA biases in English and Spanish RCs. They found that models trained on synthetic data could learn both high and LA, but models trained on real-world, multilingual data favored LA, reflecting the pattern seen in English, despite this preference not being universal across languages (see Table 1).

3 Research Questions

Considering the varied parsing outcomes across different languages, it is necessary to explore how LLMs adapt to language-specific attachment preferences. Additionally, psycholinguistic research has consistently shown that human sentence processing is deeply influenced by various linguistic factors. This leads to a broader inquiry into whether LLMs reflect patterns of sentence processing akin to those found in human linguistic behavior. Additionally, it is also important to assess whether the significance assigned to these factors differs across models. Our specific research questions are below.

- Do LLMs accurately identify relative clauses (RCs) of varying lengths across multiple languages??

- Do LLMs accurately reflect language-specific attachment preferences?

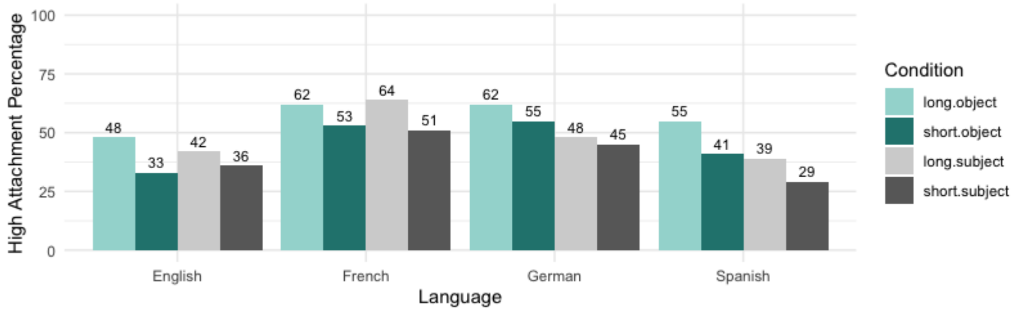

Figure 2: Human sentence processing results (Hemforth et al., 2015)

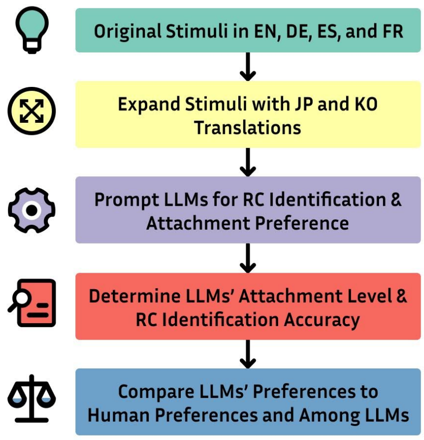

Figure 3: Overview of methodology

- Do LLMs exhibit influences in the same direction as observed in human sentence processing with regard to linguistic factors (e.g. the length of RCs and the position of the complex DP)?

Addressing these questions is crucial for understanding LLMs' processing of complex linguistic structures and for refining them to better mimic human-like capabilities in diverse language environments.

4 RC Attachment Ambiguity Resolution

To directly compare LLMs' processing results to those of humans, we replicated the experiments from Hemforth et al. (2015). Figure 3 presents the overview of this study.

4.1 Models

We evaluated five large language models (LLMs): Claude 3 Opus (Anthropic, 2024), Gemini-1.5 Pro (Team et al., 2024), GPT-3.5 (OpenAI), GPT-4o (OpenAI et al., 2024), and Llama 3 70B (Dubey et al., 2024). These include both leading proprietary models and a popular open source model.

GPT-3.5 (175B parameters) and GPT-4o (number of parameters not published), developed by OpenAI, are known for their extensive training datasets and strong multilingual performance. Claude 3 Opus (unpublished sizes) from Anthropic emphasizes reliable outputs. Gemini-1.5 Pro is a Mixture-of-Experts model from Google, and Llama 3 70B is a robust open source model.

4.2 Dataset

Our data consists of 32 sets of items in a single language, identical to those used in Hemforth et al. (2015). The experiment was conducted in six languages: English, Spanish, French, German, Japanese, and Korean. The original dataset from Hemforth et al. (2015) included translations from English to Spanish, French, and German. To extend this dataset, we obtained translations of the English stimuli into Japanese and Korean using GPT-4o (OpenAI et al., 2024), which were further refined by native speakers (details in Appendix C).

The stimuli in our experiment vary across two factors: the length of the relative clauses (RCs) (short vs. long) and the position of the complex determiner phrases (DPs) (subject vs. object). An example set of stimuli is presented in Table 2. In each language, we categorized the data into two syntactic groups: head-initial (SVO) languages—English, German, French, and Spanish—and head-final (SOV) languages—Japanese and Korean. In head-initial languages, the relative clause is postnominal, while in head-final languages, it is prenominal.

Regardless of the position of the RCs, the adjacent DP (local DP) typically serves as the LA site, while the non-adjacent DP (non-local DP) functions as the HA site. This leads to a mirror-like word order in head-initial languages (DP1 preceding DP2 within the RC) compared to head-final languages, where the RC precedes DP2 of DP1.

Position	RC length	Sentence
Subject	Short	The relative of the actor who drank hated the cameraman.
Subject	Long	The relative of the actor who too frequently drank hated the cameraman.
Object	Short	The cameraman hated the relative of the actor who drank.
Object	Long	The cameraman hated the relative of the actor who too frequently drank.

Table 2: Example set of English stimuli

Japanese and Korean, both head-final languages, are typologically similar (often grouped under the Altaic language family) and differ significantly from European languages. These typological differences can affect language model performance, as models may find it challenging to process features of head-final languages that are less familiar compared to European languages. Although there are no existing human sentence processing results for Korean and Japanese, including these languages allows us to evaluate LLMs' performance on structurally distinct languages not previously studied in Hemforth et al. (2015).

4.3 Experimental Procedure

Following Hemforth et al. (2015)'s methodology, we also conducted a comprehension task (forced-choice task). While Hemforth et al. (2015) provided specific RCs and asked participants to fill in the blank based on their interpretation, as in (3) below, we provided general instructions for the task in our experiment (4).

(3) a. The boss of the man who had a long gray beard was on vacation.

b. The ______ had a long gray beard.

(4) "Read the sentence, then 1) identify the relative clause in the sentence and 2) identify the person that the relative clause modifies. Give the correct or most likely correct answers to the two questions without commentary."

The prompt was translated into each language (see Appendix A for the full prompt texts), and the version corresponding to the sentence language was used in each case. We included RC identification (the first part of the prompt) to examine the effect of RC length on identification rates of each LLM.

5 Analysis and Results

In our analysis, we included only correct responses that accurately identified RCs. Outliers, which constituted 13.68 percent of the total data—broken down as English: 0.15%, Spanish: 4.68%, French: 2.34%, German: 3.43%, Japanese: 57.81%, and Korean: 53.75%—were excluded. Additionally, instances where the model responded with a noun other than DP1 or DP2, or declined to provide an answer for any reason, were treated as failures and removed from the dataset. Data were analyzed using mixed effects logistic regression through the lmer function from the lme4 package (Bates, 2007) in the R software 4.3.3. The main model incorporated DP position and RC length as fixed factors, with items as random factors. When constructing models, we started with the maximal random effect structure and progressively simplified it until the model converged (Barr et al., 2013). The analysis provided coefficients, standard errors, Z scores, and p-values for each fixed effect and interaction. A coefficient was considered significant at a threshold of 0.05. Note that due to the limited sample size available within each condition, we conducted our statistical analyses separately for each language, without further subdividing by model types. This approach was necessary to ensure sufficient data points for robust analysis and to mitigate issues related to model convergence.

5.1 Relative Clause Identification

RC identification results are summarized in Figure 4. Overall, the models demonstrate higher performance in head-initial languages compared to head-final languages. Specifically, Claude 3 Opus, Gemini-1.5 Pro, and Llama 3 70B show consistently high performance in English, Spanish, French, and German, with counts around 128 for each language. This consistency suggests robust training across these head-initial languages. Notably, Claude 3 Opus maintains high performance in Japanese and Korean, indicating superior training or adaptation capabilities for these left-branching Asian languages, compared to the other models.

In contrast, GPT-3.5 and GPT-4o display slightly

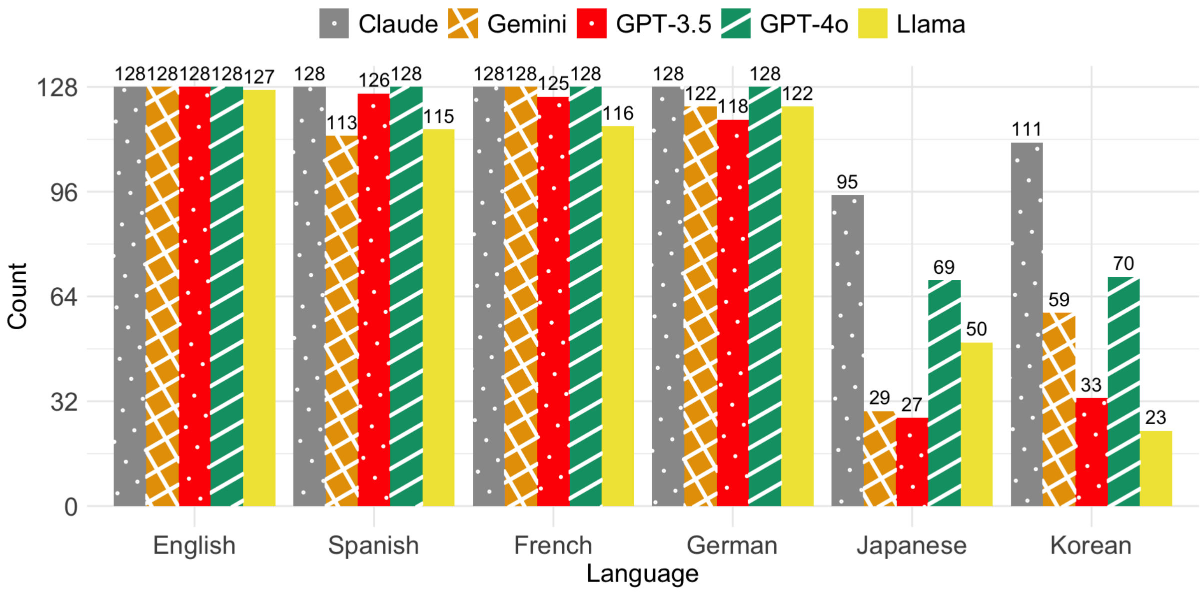

Figure 4: Models' performance on RC identification by languages: raw counts of the successful RC identification

lower performance across all languages, with a more pronounced decline for Japanese and Korean. Gemini-1.5 Pro and Llama 3 70B follow similar performance patterns across all languages, which may reflect similarities in their model architectures or training data. These findings highlight the varying generalization capabilities of each model across different linguistic contexts and suggest that some models may require further refinement, particularly in handling non-European languages.

5.2 Attachment Preferences (HA *vs.* LA)

As for the overall attachment preference, human performance indicates an LA preference in English and Spanish, and an HA preference in French, German, Japanese, and Korean. Table 3 highlights significant differences in LLMs' handling of attachment ambiguities across six languages.

In English, all models show an LA preference (scores below 30%), aligning with human preference. In Spanish, despite humans preferring LA, Gemini-1.5 Pro and GPT-3.5 show HA preferences with scores of over 80%. Claude 3 Opus shows moderate HA preference (48.03%), while Llama 3 70B and GPT-4o show LA preferences (39.13% and 21.87%). In French and German, where humans exhibit HA preferences, Claude 3 Opus, Gemini-1.5 Pro, and GPT-3.5 align with the HA preference (scores above 50%), while Llama 3 70B and GPT-4o show LA preferences.

For Japanese and Korean, which both have HA preferences in the prior human studies, models perform differently. In Japanese, most models align with the HA preference, except GPT-3.5 (33.33%). In Korean, however, all models show LA preferences (scores below 15%), in a marked divergence from humans' HA preference in Korean.

Overall, these results reveal that models exhibit different attachment preferences across languages, indicating that they process languages distinctly. However, these results do not always align with human sentence processing outcomes.

5.3 The Effect of Relative Clause Length & Syntactic Position

Figure 5 illustrates the results of further analysis concerning the effect of RC length and complex DP position on sentence structures. The statistical results are summarized in Appendix B. It is observed that models display varying preferences under different conditions across languages. Notably, when analyzing conditions involving long RCs, there is a slight increase in the preference for HA across languages. This suggests that the additional context provided by longer phrases tends to enhance the models' inclination toward HA strategies, even if it does not completely shift the overall preference in that language.

Let us now turn our attention to the effect of complex DP positions on attachment preferences across languages. In English, models predominantly exhibit LA preferences for both object and subject positions, with a slight inclination toward higher attachment in object positions, as seen in

Model	English	Spanish	French	German	Japanese	Korean
Claude 3 Opus	0.91	48.03	54.33	65.07	54.83	1.14
Gemini-1.5 Pro	22.65	88.49	81.10	93.44	51.72	12.96
GPT-3.5	27.61	80.95	79.2	69.49	33.33	12
GPT-4o	0.78	21.87	26.56	26.56	69.69	3.57
Llama 3 70B	0.83	39.13	43.96	45.9	60	0

Table 3: The high attachment answers (%) of 5 models across languages (green: HA, grey: LA)

GPT-3.5. In contrast, Spanish, French, and German generally show a stronger preference for HA in object positions, although variations exist between the models; notably, Gemini-1.5 Pro often deviates from this trend. Japanese models display mixed outcomes; for instance, GPT-4o shows a distinct preference for HA in object positions, unlike other models which do not consistently exhibit this pattern. Similar to English, Korean shows consistently LA preferences across all models and both positions.

These results show that the length of RC and the syntactic positions of the complex DP can influence attachment strategies in each model. LLMs generally exhibit similar tendencies to humans in how they handle the length of RCs and the positioning of complex DP. The models often show an increased preference for HA with longer RCs, which aligns with how humans typically process more context as a cue for attachment. However, these models may not always perfectly mimic human processing, especially across varied linguistic contexts.

While there is a general trend towards HA in longer RCs across languages, the impact of linguistic factors like RC length and DP position, indeed, varies across different language models. Some models, such as Gemini-1.5 Pro and GPT-3.5, consistently show strong preferences for HA across languages, demonstrating robust syntactic processing capabilities. In contrast, other models like GPT-4o and Llama 3 70B display more variable responses. This indicates that the interpretation of linguistic elements, such as RC length and DP position, by models is influenced by their architecture, training data, and specific training methodologies.

6 Discussion

This study holds significance in directly comparing the outcomes of LLMs on attachment ambiguity resolution with human results, as well as in analyzing the performance of each model across languages and the influence of linguistic elements on processing. The overall results show that models display varied attachment preferences across languages, suggesting distinct processing mechanisms. However, these outcomes do not consistently match human sentence processing patterns.

Among many reasons, we first speculate that such results occur because the models do not process in the given languages. Notably, in Japanese and Korean, we observed that despite the language of the input not being English, most responses were still generated in English. Thus, through the models' responses, we could confirm that especially when dealing with Asian languages, there appears to be a translation process into English.[3] This phenomenon was not observed in European languages. Our observations about internal translation are consistent with the findings of (Wendler et al., 2024), which demonstrated that in Llama-2, even during non-English tasks, the intermediate layer representations often correspond closely to English tokens. This suggests a form of internal translation even when processing inputs in other languages.

Internal machine translation often leads to errors in identifying RCs due to reliance on English—a language with different syntactic structures—to interpret syntax in Japanese and Korean. Mistranslations are likely influenced by the unique linguistic features of these languages. For instance, Japanese and Korean do not use separate relative pronouns; instead, they utilize specific morphemes to mark modifiers. These morphemes can be ambiguous and resemble other modifiers within sentences, complicating the models' ability to distinguish RCs clearly. Moreover, when translating from English back to Japanese or Korean, discrepancies occur be-

[3]This occurred most often with Korean data: Gemini-1.5 Pro included English in the response for 7 of the sentences, while Llama 3 70B responded almost entirely in English with only a few Korean phrases included. In the Japanese data, Gemini-1.5 Pro included English in 12 of the sentence responses, while Llama 3 70B had two responses that included English.

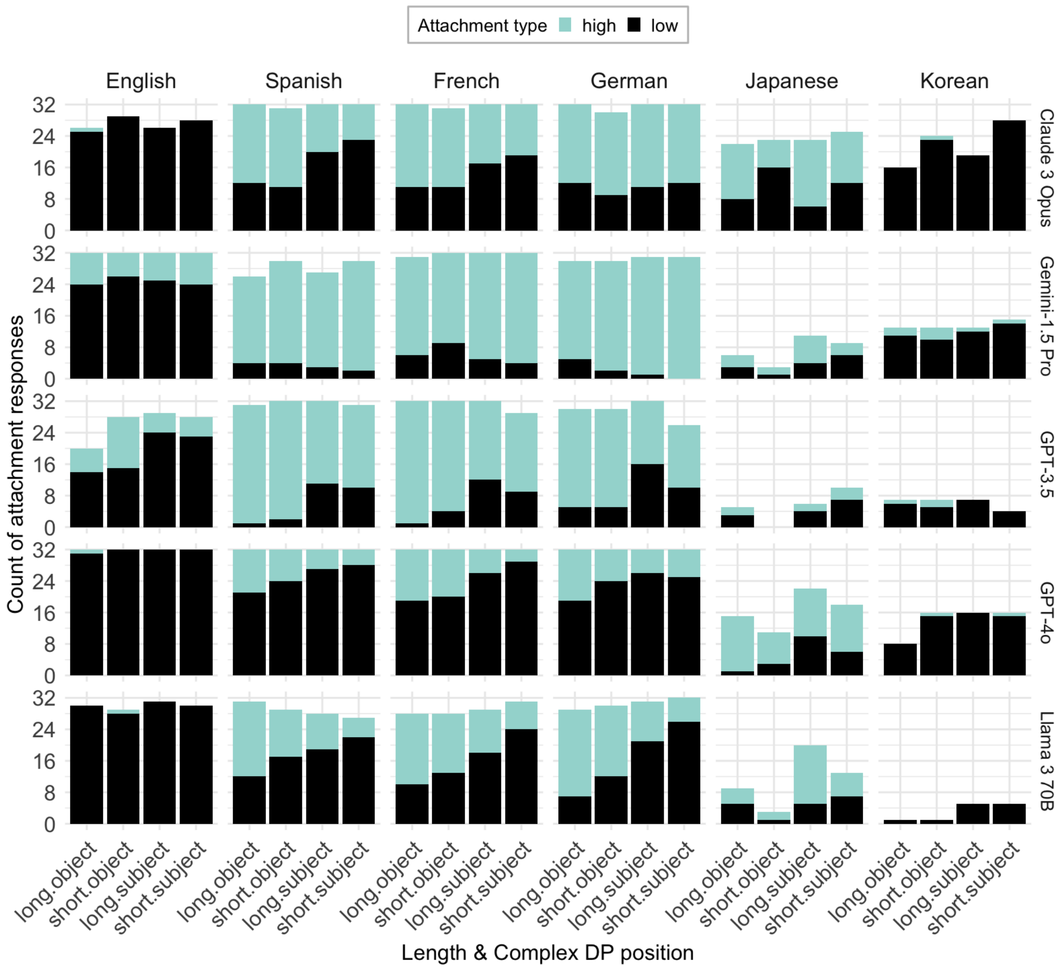

Figure 5: Distribution of attachment answers by model and language

cause the models rebuild the text based on context-heavy English inputs and learned patterns rather than the original input. This process can alter the form of RCs or introduce ambiguity with other sentence modifiers, posing significant challenges in RC identification. Observations from Gemini's and Llama 3's responses confirm that these translation errors are often linked to internal machine translation issues. This observation underscores the challenges models face when operating in languages different from their primary training language, which often leads to defaulting to English.

Interestingly, although English responses appeared in both Korean and Japanese experiments—suggestive of internal machine translation—the behaviors of LLMs in resolving RC attachment ambiguities differ markedly between these two languages. While the results in Korean exhibit a clear bias influenced by English processing patterns, such a bias is not evident in the Japanese data. According to a linguistic taxonomy (Joshi et al., 2020) which categorizes languages based on the amount of language resources available for training LLMs, all languages in our study except Korean are considered to be high resource languages, meaning that the models have access to considerable amounts of data in these languages. This disparity in resources in turn has shown to affect the downstream performance of models, with more reliable and accurate performance for higher-resource languages than for lower-resource languages (Guerreiro et al., 2023; Jin et al., 2024, a.o.).

7 Conclusion

This paper investigates how LLMs handle the understudied issue of RC attachment ambiguity, providing insights into model characteristics and their ability to mimic human-like sentence processing. The study highlights the strengths and limitations of these models in managing complex linguistic phenomena across different languages.

Acknowledgments

We are thankful to the anonymous reviewers for their helpful feedback.

References

Carlos Acuna-Farina, Isabel Fraga, Javier García-Orza, and Ana Piñeiro. 2009. Animacy in the adjunction of spanish rcs to complex nps. *European Journal of Cognitive Psychology*, 21(8):1137–1165.

Anthropic. 2024. Introducing the next generation of Claude — anthropic.com. https://www.anthropic.com/news/claude-3-family. [Accessed 30-08-2024].

Petra Augurzky. 2006. *Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing*. Ph.D. thesis, Max Planck Institute for Human Cognitive and Brain Sciences Leipzig.

Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Douglas M Bates. 2007. lme4: Linear mixed-effects models using s4 classes. *(No Title)*.

Katy Carlson, Michael Walsh Dickey, Lyn Frazier, and Charles Clifton Jr. 2009. Information structure expectations in sentence comprehension. *Quarterly Journal of Experimental Psychology*, 62(1):114–139.

Manuel Carreiras and Charles Clifton Jr. 1993. Relative clause interpretation preferences in spanish and english. *Language and speech*, 36(4):353–372.

Fernando Cuetos and Don C Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in spanish. *Cognition*, 30(1):73–105.

Forrest Davis and Marten van Schijndel. 2020. Recurrent neural network language models always learn English-like relative clause attachment. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1979–1990, Online. Association for Computational Linguistics.

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 Montgomery, 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, Ibrahim Damlaj, 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, Vítor Albiero, 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.

E Fernández. 2005. The prosody produced by spanish-english bilinguals: A preliminary investigation and implications for sentence processing. *Revista da ABRALIN*, 4(1):109–141.

Janet Dean Fodor. 1998. Learning to parse? *Journal of psycholinguistic research*, 27:285–319.

Isabel Fraga, Javier García-Orza, and Juan Carlos Acuña. 2005. La desambiguación de oraciones de relativo en gallego: Nueva evidencia de adjunción alta en lenguas romances. *Psicológica*, 26(2):243–260.

Elizabeth Gilboy, Josep-MMaria Sopena, Charles Cliftrn Jr, and Lyn Frazier. 1995. Argument structure and association preferences in spanish and english complex nps. *Cognition*, 54(2):131–167.

Nino Grillo and João Costa. 2014. A novel argument for the universality of parsing principles. *Cognition*, 133(1):156–187.

Nuno M Guerreiro, Duarte M Alves, Jonas Waldendorf, Barry Haddow, Alexandra Birch, Pierre Colombo, and André FT Martins. 2023. Hallucinations in large multilingual translation models. *Transactions of the Association for Computational Linguistics*, 11:1500–1517.

B Hemforth and L Konieczny. 2002. Where pronouns and relative clauses differ: Information structure and binding preferences. In *15th Annual CUNY Conference on Human Sentence Processing, New York, NY*.

B Hemforth, L Konieczny, and C Scheepers. 1996. Syntactic and anaphoric processes in modifier attachment. In *The 9th Annual CUNY Conference on Human Sentence Processing*, pages 21–23.

Barbara Hemforth, Susana Fernandez, Charles Clifton Jr, Lyn Frazier, Lars Konieczny, and Michael Walter. 2015. Relative clause attachment in german, english, spanish and french: Effects of position and length. *Lingua*, 166:43–64.

Yiqiao Jin, Mohit Chandra, Gaurav Verma, Yibo Hu, Munmun De Choudhury, and Srijan Kumar. 2024. Better to ask in english: Cross-lingual evaluation of large language models for healthcare queries. In *Proceedings of the ACM on Web Conference 2024*, pages 2627–2638.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Yuki Kamide and Don C Mitchell. 1997. Relative clause attachment: Nondeterminism in japanese parsing. *Journal of Psycholinguistic Research*, 26:247–254.

So Young Lee. 2021. The effect of honorific affix on pro-cessing of an attachment ambiguity. *Japanese/Korean Linguistics*, 28:1–10.

Marius Mosbach, Stefania Degaetano-Ortlieb, Marie-Pauline Krielke, Badr M. Abdullah, and Dietrich Klakow. 2020. A closer look at linguistic knowledge in masked language models: The case of relative clauses in American English. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 771–787, Barcelona, Spain (Online). International Committee on Computational Linguistics.

OpenAI. Models. https://platform.openai.com/docs/models/gpt-3-5-turbo. [Accessed 30-08-2024].

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Grusha Prasad and Tal Linzen. 2024. Spawning structural priming predictions from a cognitively motivated parser. *arXiv preprint arXiv:2403.07202*.

Grusha Prasad, Marten van Schijndel, and Tal Linzen. 2019. Using priming to uncover the organization of syntactic representations in neural language models. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 66–76, Hong Kong, China. Association for Computational Linguistics.

Amy Schafer. 1996. Focus in relative clause construal. *Language and cognitive processes*, 11(1-2):135–164.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, Soroosh Mariooryad, Yifan Ding, Xinyang Geng, Fred Alcober, Roy Frostig, Mark Omernick, Lexi Walker, Cosmin Paduraru, Christina Sorokin, Andrea Tacchetti, Colin Gaffney, Samira Daruki, Olcan Sercinoglu, Zach Gleicher, Juliette Love, Paul Voigtlaender, Rohan Jain, Gabriela Surita, Kareem Mohamed, Rory Blevins, Junwhan Ahn, Tao Zhu, Kornraphop Kawintiranon, Orhan Firat, Yiming Gu, Yujing Zhang, Matthew Rahtz, Manaal Faruqui, Natalie Clay, Justin Gilmer, JD Co-Reyes, Ivo Penchev, Rui Zhu, Nobuyuki Morioka, Kevin Hui, Krishna Haridasan, Victor Campos, Mahdis Mahdieh, Mandy Guo, Samer Hassan, Kevin Kilgour, Arpi Vezer, Heng-Tze Cheng, Raoul de Liedekerke, Siddharth Goyal, Paul Barham, DJ Strouse, Seb Noury, Jonas Adler, Mukund Sundararajan, Sharad Vikram, Dmitry Lepikhin, Michela Paganini, Xavier Garcia, Fan Yang, Dasha Valter, Maja Trebacz, Kiran Vodrahalli, Chulayuth Asawaroengchai, Roman Ring, Norbert Kalb, Livio Baldini Soares, Siddhartha Brahma, David Steiner, Tianhe Yu, Fabian Mentzer, Antoine He, Lucas Gonzalez, Bibo Xu, Raphael Lopez Kaufman, Laurent El Shafey, Junhyuk Oh, Tom Hennigan, George van den Driessche, Seth Odoom, Mario Lucic, Becca Roelofs, Sid Lall, Amit Marathe, Betty Chan, Santiago Ontanon, Luheng He, Denis Teplyashin, Jonathan Lai, Phil Crone, Bogdan Damoc, Lewis Ho, Sebastian Riedel, Karel Lenc, Chih-Kuan Yeh, Aakanksha Chowdhery, Yang Xu, Mehran Kazemi, Ehsan Amid, Anastasia Petrushkina, Kevin Swersky, Ali Khodaei, Gowoon Chen, Chris Larkin, Mario Pinto, Geng Yan, Adria Puigdomenech Badia, Piyush Patil, Steven Hansen, Dave Orr, Sebastien M. R. Arnold, Jordan Grimstad, Andrew Dai, Sholto Douglas, Rishika Sinha, Vikas Yadav, Xi Chen, Elena Gribovskaya, Jacob Austin, Jeffrey Zhao, Kaushal Patel, Paul Komarek, Sophia Austin, Sebastian Borgeaud, Linda Friso, Abhimanyu Goyal, Ben Caine, Kris Cao, Da-Woon Chung, Matthew Lamm, Gabe Barth-Maron, Thais Kagohara, Kate Olszewska, Mia Chen, Kaushik Shivakumar, Rishabh Agarwal, Harshal Godhia, Ravi Rajwar, Javier Snaider, Xerxes Dotiwalla, Yuan Liu, Aditya Barua, Victor Ungureanu, Yuan Zhang, Bat-Orgil Batsaikhan, Mateo Wirth, James Qin, Ivo Danihelka, Tulsee Doshi, Martin Chadwick, Jilin Chen, Sanil Jain, Quoc Le, Arjun Kar, Madhu Gurumurthy, Cheng Li, Ruoxin Sang, Fangyu Liu, Lampros Lamprou, Rich Munoz, Nathan Lintz, Harsh Mehta, Heidi Howard, Malcolm Reynolds, Lora Aroyo, Quan Wang, Lorenzo Blanco, Albin Cassirer, Jordan Griffith, Dipanjan Das, Stephan Lee, Jakub Sygnowski, Zach Fisher, James Besley, Richard Powell, Zafarali Ahmed, Dominik Paulus, David Reitter, Zalan Borsos, Rishabh Joshi, Aedan Pope, Steven Hand, Vittorio Selo, Vihan Jain, Nikhil Sethi, Megha Goel, Takaki Makino, Rhys May, Zhen Yang, Johan Schalkwyk, Christina Butterfield, Anja Hauth, Alex Goldin, Will Hawkins, Evan Senter, Sergey Brin, Oliver Woodman, Marvin Ritter, Eric Noland, Minh Giang, Vijay Bolina, Lisa Lee, Tim Blyth, Ian Mackinnon, Machel Reid, Obaid Sarvana, David Silver, Alexander Chen, Lily Wang, Loren Maggiore, Oscar Chang, Nithya Attaluri, Gregory Thornton, Chung-Cheng Chiu, Oskar Bunyan, Nir Levine, Timothy Chung, Evgenii Eltyshev, Xiance Si, Timothy Lillicrap, Demetra Brady, Vaibhav Aggarwal, Boxi Wu, Yuanzhong Xu, Ross McIlroy, Kartikeya Badola, Paramjit Sandhu, Erica Moreira, Wojciech Stokowiec, Ross Hemsley, Dong Li, Alex Tudor, Pranav Shyam, Elahe Rahimtoroghi, Salem Haykal, Pablo Sprechmann, Xiang Zhou, Diana Mincu, Yujia Li, Ravi Addanki, Kalpesh Krishna, Xiao Wu, Alexandre Frechette, Matan Eyal, Allan Dafoe, Dave Lacey, Jay Whang, Thi Avrahami, Ye Zhang, Emanuel Taropa, Hanzhao Lin, Daniel Toyama, Eliza Rutherford, Motoki Sano, HyunJeong Choe, Alex Tomala, Chalence Safranek-Shrader, Nora Kassner, Mantas Pajarskas, Matt Harvey, Sean Sechrist, Meire Fortunato, Christina Lyu, Gamaleldin Elsayed, Chenkai Kuang, James Lottes, Eric Chu, Chao Jia, Chih-Wei Chen, Peter Humphreys, Kate Baumli, Connie Tao, Rajkumar Samuel, Cicero Nogueira dos Santos, Anders Andreassen, Nemanja Rakićević, Dominik Grewe, Aviral Kumar, Stephanie Winkler, Jonathan Caton, Andrew Brock, Sid Dalmia, Hannah Sheahan, Iain Barr, Yingjie Miao, Paul Natsev, Jacob Devlin, Feryal Behbahani, Flavien Prost, Yanhua Sun, Artiom Myaskovsky, Thanumalayan Sankaranarayana Pillai, Dan Hurt, Angeliki Lazaridou, Xi Xiong, Ce Zheng, Fabio Pardo, Xiaowei Li, Dan Horgan, Joe Stanton, Moran Ambar, Fei Xia, Alejandro Lince, Mingqiu Wang, Basil Mustafa, Albert Webson, Hyo Lee, Rohan Anil, Martin Wicke, Timothy Dozat, Abhishek Sinha, Enrique Piqueras, Elahe Dabir, Shyam Upadhyay, Anudhyan Boral, Lisa Anne Hendricks, Corey Fry, Josip Djolonga, Yi Su, Jake Walker, Jane Labanowski, Ronny Huang, Vedant Misra, Jeremy Chen, RJ Skerry-Ryan, Avi Singh, Shruti Rijhwani, Dian Yu, Alex Castro-Ros, Beer Changpinyo, Romina Datta, Sumit Bagri, Arnar Mar Hrafnkelsson, Marcello Maggioni, Daniel Zheng, Yury Sulsky, Shaobo Hou, Tom Le Paine, Antoine Yang, Jason Riesa, Dominika Rogozinska, Dror Marcus, Dalia El Badawy, Qiao Zhang, Luyu Wang, Helen Miller, Jeremy Greer, Lars Lowe Sjos, Azade Nova, Heiga Zen, Rahma Chaabouni, Mihaela Rosca, Jiepu Jiang, Charlie Chen, Ruibo Liu, Tara Sainath, Maxim Krikun, Alex Polozov, Jean-Baptiste Lespiau, Josh Newlan, Zeyncep Cankara, Soo Kwak, Yunhan Xu, Phil Chen, Andy Coenen, Clemens Meyer, Katerina

Tsihlas, Ada Ma, Juraj Gottweis, Jinwei Xing, Chenjie Gu, Jin Miao, Christian Frank, Zeynep Cankara, Sanjay Ganapathy, Ishita Dasgupta, Steph Hughes-Fitt, Heng Chen, David Reid, Keran Rong, Hongmin Fan, Joost van Amersfoort, Vincent Zhuang, Aaron Cohen, Shixiang Shane Gu, Anhad Mohananey, Anastasija Ilic, Taylor Tobin, John Wieting, Anna Bortsova, Phoebe Thacker, Emma Wang, Emily Caveness, Justin Chiu, Eren Sezener, Alex Kaskasoli, Steven Baker, Katie Millican, Mohamed Elhawaty, Kostas Aisopos, Carl Lebsack, Nathan Byrd, Hanjun Dai, Wenhao Jia, Matthew Wiethoff, Elnaz Davoodi, Albert Weston, Lakshman Yagati, Arun Ahuja, Isabel Gao, Golan Pundak, Susan Zhang, Michael Azzam, Khe Chai Sim, Sergi Caelles, James Keeling, Abhanshu Sharma, Andy Swing, YaGuang Li, Chenxi Liu, Carrie Grimes Bostock, Yamini Bansal, Zachary Nado, Ankesh Anand, Josh Lipschultz, Abhijit Karmarkar, Lev Proleev, Abe Ittycheriah, Soheil Hassas Yeganeh, George Polovets, Aleksandra Faust, Jiao Sun, Alban Rrustemi, Pen Li, Rakesh Shivanna, Jeremiah Liu, Chris Welty, Federico Lebron, Anirudh Baddepudi, Sebastian Krause, Emilio Parisotto, Radu Soricut, Zheng Xu, Dawn Bloxwich, Melvin Johnson, Behnam Neyshabur, Justin Mao-Jones, Renshen Wang, Vinay Ramasesh, Zaheer Abbas, Arthur Guez, Constant Segal, Duc Dung Nguyen, James Svensson, Le Hou, Sarah York, Kieran Milan, Sophie Bridgers, Wiktor Gworek, Marco Tagliasacchi, James Lee-Thorp, Michael Chang, Alexey Guseynov, Ale Jakse Hartman, Michael Kwong, Ruizhe Zhao, Sheleem Kashem, Elizabeth Cole, Antoine Miech, Richard Tanburn, Mary Phuong, Filip Pavetic, Sebastien Cevey, Ramona Comanescu, Richard Ives, Sherry Yang, Cosmo Du, Bo Li, Zizhao Zhang, Mariko Iinuma, Clara Huiyi Hu, Aurko Roy, Shaan Bijwadia, Zhenkai Zhu, Danilo Martins, Rachel Saputro, Anita Gergely, Steven Zheng, Dawei Jia, Ioannis Antonoglou, Adam Sadovsky, Shane Gu, Yingying Bi, Alek Andreev, Sina Samangooei, Mina Khan, Tomas Kocisky, Angelos Filos, Chintu Kumar, Colton Bishop, Adams Yu, Sarah Hodkinson, Sid Mittal, Premal Shah, Alexandre Moufarek, Yong Cheng, Adam Bloniarz, Jaehoon Lee, Pedram Pejman, Paul Michel, Stephen Spencer, Vladimir Feinberg, Xuehan Xiong, Nikolay Savinov, Charlotte Smith, Siamak Shakeri, Dustin Tran, Mary Chesus, Bernd Bohnet, George Tucker, Tamara von Glehn, Carrie Muir, Yiran Mao, Hideto Kazawa, Ambrose Slone, Kedar Soparkar, Disha Shrivastava, James Cobon-Kerr, Michael Sharman, Jay Pavagadhi, Carlos Araya, Karolis Misiunas, Nimesh Ghelani, Michael Laskin, David Barker, Qiujia Li, Anton Briukhov, Neil Houlsby, Mia Glaese, Balaji Lakshminarayanan, Nathan Schucher, Yunhao Tang, Eli Collins, Hyeontaek Lim, Fangxiaoyu Feng, Adria Recasens, Guangda Lai, Alberto Magni, Nicola De Cao, Aditya Siddhant, Zoe Ashwood, Jordi Orbay, Mostafa Dehghani, Jenny Brennan, Yifan He, Kelvin Xu, Yang Gao, Carl Saroufim, James Molloy, Xinyi Wu, Seb Arnold, Solomon Chang, Julian Schrittwieser, Elena Buchatskaya, Soroush Radpour, Martin Polacek, Skye Giordano, Ankur Bapna, Simon Tokumine, Vincent Hellendoorn, Thibault Sottiaux, Sarah Cogan, Aliaksei Severyn, Mohammad Saleh, Shantanu Thakoor, Laurent Shefey, Siyuan Qiao, Meenu Gaba, Shuo yiin Chang, Craig Swanson, Biao Zhang, Benjamin Lee, Paul Kishan Rubenstein, Gan Song, Tom Kwiatkowski, Anna Koop, Ajay Kannan, David Kao, Parker Schuh, Axel Stjerngren, Golnaz Ghiasi, Gena Gibson, Luke Vilnis, Ye Yuan, Felipe Tiengo Ferreira, Aishwarya Kamath, Ted Klimenko, Ken Franko, Kefan Xiao, Indro Bhattacharya, Miteyan Patel, Rui Wang, Alex Morris, Robin Strudel, Vivek Sharma, Peter Choy, Sayed Hadi Hashemi, Jessica Landon, Mara Finkelstein, Priya Jhakra, Justin Frye, Megan Barnes, Matthew Mauger, Dennis Daun, Khuslen Baatarsukh, Matthew Tung, Wael Farhan, Henryk Michalewski, Fabio Viola, Felix de Chaumont Quitry, Charline Le Lan, Tom Hudson, Qingze Wang, Felix Fischer, Ivy Zheng, Elspeth White, Anca Dragan, Jean baptiste Alayrac, Eric Ni, Alexander Pritzel, Adam Iwanicki, Michael Isard, Anna Bulanova, Lukas Zilka, Ethan Dyer, Devendra Sachan, Srivatsan Srinivasan, Hannah Muckenhirn, Honglong Cai, Amol Mandhane, Mukarram Tariq, Jack W. Rae, Gary Wang, Kareem Ayoub, Nicholas FitzGerald, Yao Zhao, Woohyun Han, Chris Alberti, Dan Garrette, Kashyap Krishnakumar, Mai Gimenez, Anselm Levskaya, Daniel Sohn, Josip Matak, Inaki Iturrate, Michael B. Chang, Jackie Xiang, Yuan Cao, Nishant Ranka, Geoff Brown, Adrian Hutter, Vahab Mirrokni, Nanxin Chen, Kaisheng Yao, Zoltan Egyed, Francois Galilee, Tyler Liechty, Praveen Kallakuri, Evan Palmer, Sanjay Ghemawat, Jasmine Liu, David Tao, Chloe Thornton, Tim Green, Mimi Jasarevic, Sharon Lin, Victor Cotruta, Yi-Xuan Tan, Noah Fiedel, Hongkun Yu, Ed Chi, Alexander Neitz, Jens Heitkaemper, Anu Sinha, Denny Zhou, Yi Sun, Charbel Kaed, Brice Hulse, Swaroop Mishra, Maria Georgaki, Sneha Kudugunta, Clement Farabet, Izhak Shafran, Daniel Vlasic, Anton Tsitsulin, Rajagopal Ananthanarayanan, Alen Carin, Guolong Su, Pei Sun, Shashank V, Gabriel Carvajal, Josef Broder, Iulia Comsa, Alena Repina, William Wong, Warren Weilun Chen, Peter Hawkins, Egor Filonov, Lucia Loher, Christoph Hirnschall, Weiyi Wang, Jingchen Ye, Andrea Burns, Hardie Cate, Diana Gage Wright, Federico Piccinini, Lei Zhang, Chu-Cheng Lin, Ionel Gog, Yana Kulizhskaya, Ashwin Sreevatsa, Shuang Song, Luis C. Cobo, Anand Iyer, Chetan Tekur, Guillermo Garrido, Zhuyun Xiao, Rupert Kemp, Huaixiu Steven Zheng, Hui Li, Ananth Agarwal, Christel Ngani, Kati Goshvadi, Rebeca Santamaria-Fernandez, Wojciech Fica, Xinyun Chen, Chris Gorgolewski, Sean Sun, Roopal Garg, Xinyu Ye, S. M. Ali Eslami, Nan Hua, Jon Simon, Pratik Joshi, Yelin Kim, Ian Tenney, Sahitya Potluri, Lam Nguyen Thiet, Quan Yuan, Florian Luisier, Alexandra Chronopoulou, Salvatore Scellato, Praveen Srinivasan, Minmin Chen, Vinod Koverkathu, Valentin Dalibard, Yaming Xu, Brennan Saeta, Keith Anderson, Thibault Sellam, Nick Fernando, Fantine Huot, Junehyuk Jung, Mani Varadarajan, Michael Quinn, Amit Raul, Maigo Le, Ruslan Habalov, Jon Clark, Komal Jalan, Kalesha Bullard, Achintya Singhal, Thang Luong, Boyu Wang, Sujeevan Rajayogam, Julian Eisenschlos,

Johnson Jia, Daniel Finchelstein, Alex Yakubovich, Daniel Balle, Michael Fink, Sameer Agarwal, Jing Li, Dj Dvijotham, Shalini Pal, Kai Kang, Jaclyn Konzelmann, Jennifer Beattie, Olivier Dousse, Diane Wu, Remi Crocker, Chen Elkind, Siddhartha Reddy Jonnalagadda, Jong Lee, Dan Holtmann-Rice, Krystal Kallarackal, Rosanne Liu, Denis Vnukov, Neera Vats, Luca Invernizzi, Mohsen Jafari, Huanjie Zhou, Lilly Taylor, Jennifer Prendki, Marcus Wu, Tom Eccles, Tianqi Liu, Kavya Kopparapu, Francoise Beaufays, Christof Angermueller, Andreea Marzoca, Shourya Sarcar, Hilal Dib, Jeff Stanway, Frank Perbet, Nejc Trdin, Rachel Sterneck, Andrey Khorlin, Dinghua Li, Xihui Wu, Sonam Goenka, David Madras, Sasha Goldshtein, Willi Gierke, Tong Zhou, Yaxin Liu, Yannie Liang, Anais White, Yunjie Li, Shreya Singh, Sanaz Bahargam, Mark Epstein, Sujoy Basu, Li Lao, Adnan Ozturel, Carl Crous, Alex Zhai, Han Lu, Zora Tung, Neeraj Gaur, Alanna Walton, Lucas Dixon, Ming Zhang, Amir Globerson, Grant Uy, Andrew Bolt, Olivia Wiles, Milad Nasr, Ilia Shumailov, Marco Selvi, Francesco Piccinno, Ricardo Aguilar, Sara McCarthy, Misha Khalman, Mrinal Shukla, Vlado Galic, John Carpenter, Kevin Villela, Haibin Zhang, Harry Richardson, James Martens, Matko Bosnjak, Shreyas Rammohan Belle, Jeff Seibert, Mahmoud Alnahlawi, Brian McWilliams, Sankalp Singh, Annie Louis, Wen Ding, Dan Popovici, Lenin Simicich, Laura Knight, Pulkit Mehta, Nishesh Gupta, Chongyang Shi, Saaber Fatehi, Jovana Mitrovic, Alex Grills, Joseph Pagadora, Dessie Petrova, Danielle Eisenbud, Zhishuai Zhang, Damion Yates, Bhavishya Mittal, Nilesh Tripuraneni, Yannis Assael, Thomas Brovelli, Prateek Jain, Mihajlo Velimirovic, Canfer Akbulut, Jiaqi Mu, Wolfgang Macherey, Ravin Kumar, Jun Xu, Haroon Qureshi, Gheorghe Comanici, Jeremy Wiesner, Zhitao Gong, Anton Ruddock, Matthias Bauer, Nick Felt, Anirudh GP, Anurag Arnab, Dustin Zelle, Jonas Rothfuss, Bill Rosgen, Ashish Shenoy, Bryan Seybold, Xinjian Li, Jayaram Mudigonda, Goker Erdogan, Jiawei Xia, Jiri Simsa, Andrea Michi, Yi Yao, Christopher Yew, Steven Kan, Isaac Caswell, Carey Radebaugh, Andre Elisseeff, Pedro Valenzuela, Kay McKinney, Kim Paterson, Albert Cui, Eri Latorre-Chimoto, Solomon Kim, William Zeng, Ken Durden, Priya Ponnapalli, Tiberiu Sosea, Christopher A. Choquette-Choo, James Manyika, Brona Robenek, Harsha Vashisht, Sebastien Pereira, Hoi Lam, Marko Velic, Denese Owusu-Afriyie, Katherine Lee, Tolga Bolukbasi, Alicia Parrish, Shawn Lu, Jane Park, Balaji Venkatraman, Alice Talbert, Lambert Rosique, Yuchung Cheng, Andrei Sozanschi, Adam Paszke, Praveen Kumar, Jessica Austin, Lu Li, Khalid Salama, Wooyeol Kim, Nandita Dukkipati, Anthony Baryshnikov, Christos Kaplanis, XiangHai Sheng, Yuri Chervonyi, Caglar Unlu, Diego de Las Casas, Harry Askham, Kathryn Tunyasuvunakool, Felix Gimeno, Siim Poder, Chester Kwak, Matt Miecnikowski, Vahab Mirrokni, Alek Dimitriev, Aaron Parisi, Dangyi Liu, Tomy Tsai, Toby Shevlane, Christina Kouridi, Drew Garmon, Adrian Goedeckemeyer, Adam R. Brown, Anitha Vijayakumar, Ali Elqursh, Sadegh Jazayeri, Jin Huang, Sara Mc Carthy, Jay Hoover, Lucy Kim, Sandeep Kumar, Wei Chen, Courtney Biles, Garrett Bingham, Evan Rosen, Lisa Wang, Qijun Tan, David Engel, Francesco Pongetti, Dario de Cesare, Dongseong Hwang, Lily Yu, Jennifer Pullman, Srini Narayanan, Kyle Levin, Siddharth Gopal, Megan Li, Asaf Aharoni, Trieu Trinh, Jessica Lo, Norman Casagrande, Roopali Vij, Loic Matthey, Bramandia Ramadhana, Austin Matthews, CJ Carey, Matthew Johnson, Kremena Goranova, Rohin Shah, Shereen Ashraf, Kingshuk Dasgupta, Rasmus Larsen, Yicheng Wang, Manish Reddy Vuyyuru, Chong Jiang, Joana Ijazi, Kazuki Osawa, Celine Smith, Ramya Sree Boppana, Taylan Bilal, Yuma Koizumi, Ying Xu, Yasemin Altun, Nir Shabat, Ben Bariach, Alex Korchemniy, Kiam Choo, Olaf Ronneberger, Chimezie Iwuanyanwu, Shubin Zhao, David Soergel, Cho-Jui Hsieh, Irene Cai, Shariq Iqbal, Martin Sundermeyer, Zhe Chen, Elie Bursztein, Chaitanya Malaviya, Fadi Biadsy, Prakash Shroff, Inderjit Dhillon, Tejasi Latkar, Chris Dyer, Hannah Forbes, Massimo Nicosia, Vitaly Nikolaev, Somer Greene, Marin Georgiev, Pidong Wang, Nina Martin, Hanie Sedghi, John Zhang, Praseem Banzal, Doug Fritz, Vikram Rao, Xuezhi Wang, Jiageng Zhang, Viorica Patraucean, Dayou Du, Igor Mordatch, Ivan Jurin, Lewis Liu, Ayush Dubey, Abhi Mohan, Janek Nowakowski, Vlad-Doru Ion, Nan Wei, Reiko Tojo, Maria Abi Raad, Drew A. Hudson, Vaishakh Keshava, Shubham Agrawal, Kevin Ramirez, Zhichun Wu, Hoang Nguyen, Ji Liu, Madhavi Sewak, Bryce Petrini, DongHyun Choi, Ivan Philips, Ziyue Wang, Ioana Bica, Ankush Garg, Jarek Wilkiewicz, Priyanka Agrawal, Xiaowei Li, Danhao Guo, Emily Xue, Naseer Shaik, Andrew Leach, Sadh MNM Khan, Julia Wiesinger, Sammy Jerome, Abhishek Chakladar, Alek Wenjiao Wang, Tina Ornduff, Folake Abu, Alireza Ghaffarkhah, Marcus Wainwright, Mario Cortes, Frederick Liu, Joshua Maynez, Andreas Terzis, Pouya Samangouei, Riham Mansour, Tomasz Kępa, François-Xavier Aubet, Anton Algymr, Dan Banica, Agoston Weisz, Andras Orban, Alexandre Senges, Ewa Andrejczuk, Mark Geller, Niccolo Dal Santo, Valentin Anklin, Majd Al Merey, Martin Baeuml, Trevor Strohman, Junwen Bai, Slav Petrov, Yonghui Wu, Demis Hassabis, Koray Kavukcuoglu, Jeffrey Dean, and Oriol Vinyals. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *Preprint*, arXiv:2403.05530.

Maria Tikhonova, Vladislav Mikhailov, Dina Pisarevskaya, Valentin Malykh, and Tatiana Shavrina. 2023. Ad astra or astray: Exploring linguistic knowledge of multilingual bert through nli task. *Natural Language Engineering*, 29(3):554–583.

Alex Warstadt and Samuel R Bowman. 2019. Linguistic analysis of pretrained sentence encoders with acceptability judgments. *arXiv preprint arXiv:1901.03438*.

Chris Wendler, Veniamin Veselovsky, Giovanni Monea, and Robert West. 2024. Do llamas work in English? on the latent language of multilingual transformers. In *Proceedings of the 62nd Annual Meeting of the*

Association for Computational Linguistics (Volume 1: Long Papers), pages 15366–15394, Bangkok, Thailand. Association for Computational Linguistics.

Yida Xin, Henry Lieberman, and Peter Chin. 2021. Revisiting the prepositional-phrase attachment problem using explicit commonsense knowledge. *arXiv preprint arXiv:2102.00924*.

Fan Yin, Zhouxing Shi, Cho-Jui Hsieh, and Kai-Wei Chang. 2021. On the sensitivity and stability of model interpretations in nlp. *arXiv preprint arXiv:2104.08782*.

A Prompts

The following are the prompts used for each language.

1. Read the sentence, then 1) identify the relative clause in the sentence and 2) identify the person that the relative clause modifies. Give the correct or most likely correct answers to the two questions without commentary. (EN)

2. Lea la frase, luego 1) identifique la cláusula relativa en la frase y 2) identifique la persona que la cláusula relativa modifica. Dé las respuestas correctas o más probables a las dos preguntas sin comentarios. (ES)

3. Lesen Sie den Satz, dann 1) identifizieren Sie den Relativsatz im Satz und 2) bestimmen Sie die Person, die der Relativsatz modifiziert. Geben Sie die korrekten oder wahrscheinlich korrekten Antworten auf die zwei Fragen ohne Kommentar. (DE)

4. Lisez la phrase, puis 1) identifiez la proposition relative dans la phrase et 2) identifiez la personne que la proposition relative modifie. Donnez les réponses correctes ou les plus probables aux deux questions sans commentaire. (FR)

5. 문장을 읽고, 1) 문장에서 관계절을 찾아내고 2) 그 관계절이 수정하는 사람을 식별하세요. 두 질문에 대한 정확하거나 가장 가능성 높은 답변을 논평 없이 제공하세요. (KO)

6. 文をんでから`1) 文中の係節を特定し`2) 係節が修飾している人物を特定してください°コメントなしで`2つの質問にする正しいまたは最も正しいと思われる答えを示してください° (JP)

B Statistical Analysis

The following tables summarize the statistical analysis.

C Translations

The Japanese and Korean datasets were automatically translated from the English language dataset using GPT-4o. The Korean translation was verified by a native speaker and the Japanese translation was verified by two professional translators.

Table 4: English: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	-2.50516	0.41247	-6.074	1.25e-09 ***
Length: short	0.18686	0.38826	0.481	0.630
Position: subject	-0.46795	0.42950	-1.090	0.276
Length: short × Position: subject	-0.08609	0.59030	-0.146	0.884

Table 5: Spanish: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	0.8593	0.2663	3.227	0.001253 **
Length: short	-0.2848	0.2632	-1.082	0.279204
Position: subject	-1.0184	0.2631	-3.871	0.000108 ***
Length: short × Position: subject	0.1490	0.3653	0.408	0.683401

Table 6: French: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	1.1086	0.3372	3.288	0.00101 **
Length: short	-0.4099	0.2832	-1.448	0.14774
Position: subject	-1.1428	0.2805	-4.074	4.63e-05 ***
Length: short × Position: subject	0.1453	0.3866	0.376	0.70696

Table 7: German: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	0.9253	0.2683	3.448	0.000564 ***
Length: short	-0.2076	0.2690	-0.772	0.440153
Position: subject	-0.8416	0.2620	-3.212	0.001317 **
Length: short × Position: subject	0.2126	0.3675	0.579	0.562885

Table 8: Japanese: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	0.8640	0.5221	1.655	0.09795 .
Length: short	-1.5506	0.5752	-2.696	0.00703 **
Position: subject	0.3748	0.4751	0.789	0.43020
Length: short × Position: subject	0.1031	0.6925	0.149	0.88167

Table 9: Korean: Statistical Analysis Results

Term	Estimate	Std. Error	z value	Pr(>lzl)
Intercept	-6.9238	2.4147	-2.867	0.00414 **
Length: short	0.8328	1.0717	0.777	0.43710
Position: subject	-1.1847	1.5938	-0.743	0.45728
Length: short × Position: subject	-1.2155	1.9252	-0.631	0.52781