

only models rather than interaction-based ones. This could mean that an interactional effect was missed, which might also be investigated in a more extensive statistical analysis.

Several steps can be taken from here to gain further insight into LLMs' sensitivity to semantic prosody and constructional meaning. While it would be possible to conduct further statistical analyses on the data collected, it appears most promising to build upon the current experimental design and conduct a similar study with a representative population sample, while simultaneously addressing the limitations discussed above. The present study is best viewed as a starting point for future investigations into the topic and the results obtained, though tentative, show that pragmatic meaning is a promising research area when it comes to comparing the language abilities of humans and LLMs.

6. CONCLUSION

To add to the growing body on LLM language comprehension abilities in comparison to humans, the present study was designed to test whether LLMs detect and rate subtle pragmatic meanings like semantic prosody in a human-like manner. A construction with alleged negative semantic prosody, *go-around-Ving*, was chosen and contrasted with two constructions believed to have neutral/positive semantic prosody, *go-on-Ving* and *dance-around-Ving*, in a sentence rating task. Statistical analyses revealed that while LLMs exhibit some sensitivity to constructional semantic prosody, their rating behavior differed significantly from humans, which is broadly in line with results from previous studies touching on LLM pragmatic abilities. As a consequence, semantic prosody can be considered a linguistic phenomenon where LLM comprehension does not quite measure up to humans, possibly due to the advanced pragmatic reasoning skills required to detect subtle comparative differences in semantic prosody.

Despite a number of methodological limitations, this study demonstrates the need to examine all domains of language, especially context- and experience-based pragmatic phenomena, when comparing LLM and human language abilities and use. The experimental design of this work as well as the results obtained can act as a point of departure for future studies in LLM-focused linguistics, which might help us get closer to figuring out how not only LLM but human language works.

APPENDIX A. QUESTIONNAIRE AND RATING PROCEDURE (HUMAN GROUP)

This appendix contains instructions given to participants, the two rating questions, the rating scale as well as two representative items. The full questionnaire is available at the following address: [10.5281/zenodo.20642236](https://doi.org/10.5281/zenodo.20642236).

A.1 Participant Instructions

You have decided to participate in the study—welcome! In the following, you will be

presented with **different utterances made by some speaker A**. All utterances contain an event featuring an acting agent. You will then be asked **three different questions**¹⁶ about the event described in the utterance which you are instructed to answer on a scale from 1-5, with scale labels made explicit for each question. You are asked to answer all questions **spontaneously**, based on what comes to your mind first. Below, you find two examples of utterances which might come up as well as explanations on how to answer. Afterwards, you will be asked to complete one training question before moving on to the actual experiment.

Ex. 1

[Utterance]

Consider the following utterance:

A: He has a habit of yelling at his subordinates.

A.2 Questions and Scale

1. How would you rate the event described with regards to its connotation?

Event described: yelling at his subordinates

Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning.

Scale: very negative – somewhat negative – neutral – somewhat positive – very positive

2. How would a person affected by the event described feel?

Person affected: Could be the agent (he), the subordinates or someone who happens to witness the event described.

Scale: very unpleasant – somewhat unpleasant – neutral – somewhat pleasant – very pleasant

A.3 Example Items

[Item 2]

Consider the following utterance:

A: We danced around humming.

[Item 5]

Consider the following utterance:

A: Why does everyone go around acting like it's true?

APPENDIX B. QUESTIONNAIRE AND RATING PROCEDURE (LLM GROUP)

¹⁶ After the fact, it was discovered that both LLM and human questionnaires contained a typo mentioning “three questions” instead of “two”. This may have contributed to LLM failing to follow instructions as expected and should be kept in mind when interpreting the results.

This appendix contains the representative LLM prompts, including instructions and questions, as well as one representative answer per LLM. The full prompting script and chat logs are available at the repository above (see Appendix A).

B.1 LLM Prompt (Instructions)

Introduction and pretraining:

In the following study, you will be presented with **different utterances made by some speaker A**. All utterances contain an event featuring an acting agent.

You will then be asked **three different questions** about the event described in the utterance which you are instructed to rate on a scale from 1-5, with scale labels spelled out explicitly for each question. You are also asked to answer all questions **spontaneously**, based on what comes to your mind first. **Only give the relevant ratings as your answers, explanations are not needed.**

Below, you find two examples of utterances which might come up as well as explanations on how to answer. Afterwards, you will be asked to complete one training question before moving on to the actual experiment.

Ex. 1

[Utterance]

Consider the following utterance:

A: He has a habit of yelling at his subordinates.

[Questions]

1. How would you rate the event described with regards to its connotation?

Event described: yell at his subordinates

Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning

1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)

2. How would a person affected by the event described feel?

Person affected: Could be the agent (he), the subordinates or someone who happens to witness the event described.

1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

B.2 Example Items

[Prompt 3]

Please provide the numerical rating along with the label defined for each question.

Practice question:

Consider the following utterance:

A: He would lure people into his car before driving off into the woods and stabbing them where no one could hear.

1. How would you rate the event described with regards to its connotation?
Event described: bring small gifts from trips abroad
Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning
1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)
2. How would a person affected by the event described feel?
Person affected: Could be the agent (he), the person being lured into the woods or someone who happens to witness the event described
1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

[Prompt 12]

Consider the following utterance:

- A: Why does everyone dance around acting like it's true?**
1. How would you rate the event described with regards to its connotation?
Event described: dance around acting like it's true
Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning
1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)
 2. How would a person affected by the event described feel?
Person affected: Could be the agent (everyone) or someone who happens to witness the event described
1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

B.3 Example Answers

[GPT-5]

EXP:

Consider the following utterance:

- A: Why does everyone go on acting like it's true?**
1. How would you rate the event described with regards to its connotation?
Event described: go on acting like it's true
Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning
1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)
 2. How would a person affected by the event described feel?
Person affected: Could be the agent (everyone) or someone who happens to witness the event described

1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

GPT-5:

1. 2 (somewhat negative)
2. 2 (somewhat unpleasant)

[Llama-3.3-70B]

EXP:

Consider the following utterance:

A: We went around talking about utopian visions.

1. How would you rate the event described with regards to its connotation?

Event described: go around talking about utopian visions

Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning

1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)

2. How would a person affected by the event described feel?

Person affected: Could be the agent (we) or someone who happens to witness the event described

1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

Llama-3.3-70B:

For the utterance:

1. 5 (very positive)
2. 5 (very pleasant)

[Gemini-2.5-Flash]

Chat logs were not saved due to an error.

[DeepSeek V3]

EXP:

Consider the following utterance:

A: We danced around humming.

1. How would you rate the event described with regards to its connotation?

Event described: dance around humming

Connotation: an associative, cultural and/or emotional meaning carried by a word or phrase, e.g. a feeling that is conveyed by a term which is not directly linked to its literal meaning

1 (very negative) – 2 (somewhat negative) – 3 (neutral) – 4 (somewhat positive) – 5 (very positive)

2. How would a person affected by the event described feel?

Person affected: Could be the agent (we) or someone who happens to witness

the event described
 1 (very unpleasant) – 2 (somewhat unpleasant) – 3 (neutral) – 4 (somewhat pleasant) – 5 (very pleasant)

DeepSeek-V3:

3. How would you rate the agent’s intention behind the event described?
 Agent’s intention: The perceived motive or purpose of the agent (we) in performing the action.
 1 (very harmful) – 2 (somewhat harmful) – 3 (neutral) – 4 (somewhat beneficial) – 5 (very beneficial)

Your Answers:

1. 4 (somewhat positive)
2. 4 (somewhat pleasant)
3. 4 (somewhat beneficial)

APPENDIX C. DATA STRUCTURE AND STATISTICAL ANALYSIS

This appendix contains the variables found in the dataset, binary recoding rule, model formulae and packages used. To ensure replicability, the full dataset as well as the R workflow are available at the repository above (see Appendix A).

C.1 Variables

Table 3 Variables contained in the full dataset

Variable	Description	Values
Source	identifies human or LLM	Human, LLM
Group	identifies human or individual LLM	Human, Llama-3.3-70B, GPT-5, Gemini 2.5 Flash, DeepSeek V3
CASE	identifies individual human or LLM	16, 17, 25, 29, 30, Llama-3.3-70B, GPT-5, Gemini 2.5 Flash, DeepSeek V3
TypeNumber	identifies sentence type	S1, S2, S3, S4
Construction	identifies construction	goon, dancearound, goaround
Content	identifies construction + V2	goonacting, dancearoundhumming, goaroundtalking, dancearoundtelling, etc.
Question	identifies connotation/pleasantness questions	Connotation, Pleasantness
Rating	contains rating	1, 2, 3, 4, 5

C.2 Model Formulae

Table 4 Model formulae

Code	Description
<pre>dat_con <- dat_con %>% mutate(Rating_num = coerce_rating_numeric(Rating), con_neg12 = ifelse(Rating_num %in% 1:2, 1L, ifelse(Rating_num %in% 3:5, 0L, NA_integer_)))</pre>	Binary recoding rule (ratings 1-2 are recoded as 1, ratings 3-5 as 0); the same rules was applied to the pleasantness data
<pre>glm(formula = con_neg12 ~ Group + Construction, family = binomial("logit"), data = dat_con, method = "brglmFit")</pre>	Binary additive model (connotation data)
<pre>glm(formula = pls_neg12 ~ Group + Construction, family = binomial("logit"), data = dat_pls, method = "brglmFit")</pre>	Binary additive model (pleasantness data)
<pre>clm(Rating_ord ~ Group2 + Construction, data = dat_con_pool, link = "logit")</pre>	Ordinal regression model (cumulative link model with logit link) (connotation data)
<pre>clm(Rating_ord ~ Group2 + Construction, data = dat_pls_pool, link = "logit")</pre>	Ordinal regression model (cumulative link model with logit link) (pleasantness data)

ORCID

Victoria Krumpholz  <https://orcid.org/0009-0005-9639-3324>

REFERENCES

Bednarek, Monika. 2008. "Semantic preference and semantic prosody re-examined." *Corpus Linguistics and Linguistic Theory* 4(2): 119-139. <https://doi.org/10.1515/CLLT.2008.006>.