John David Storment

Pictorial constituents & the metalinguistic performance of LLMs

Abstract

In this paper I show that, although ChatGPT (GPT-4o) can provide accurate linguistic acceptability judgments for many types of sentences (Cai, Duan, Haslett, Wang, & Pickering, 2024; Collins, 2024a, 2024b; Ortega-Martín et al., 2023; Wang et al., 2023), it does not give accurate grammaticality judgments for sentences that contain pro-text emojis, which are emojis that appear in a written utterance as morphosyntactic constituents (Cohn, Engelen, & Schilperoord, 2019; Pierini, 2021; Storment, 2024; Tieu, Qiu, Puvipalan, & Pasternak, 2025, a.o.). I demonstrate this with three distinct experiments performed on GPT-4o using both English and Spanish data. This work builds on prior research that shows that the combinatorics of pro-text emojis are sensitive to the morphosyntactic constraints of the language in which the emojis appear, and it connects the poor performance of GPT-4o in this respect to two factors: (i) the fact that, while LLMs are able to make some generalizations of syntactic structural dependencies, their mechanisms for making such generalizations are not derived in the same way that human syntactic structures are (Contreras Kallens, Kristensen-McLachlan, & Christiansen, 2023; Hale & Stanojević, 2024; Kennedy, 2025; Linzen & Baroni, 2021; Manova, 2024a, 2024b; Zhong, Ding, Liu, Du, & Tao, 2023, a.o.), and (ii) the fact that LLMs lack the means of directly processing iconic and pictorial content in the same way that human cognition allows for. I also consider the possibility that the relevant data are poorly attested in the model's training parameters. This paper establishes a precedent for the research of the intersection of generative AI and utterances that contain pictorial elements as morphosyntactic constituents.

1 Introduction

ChatGPT is very good at giving human-like acceptability judgments for many English sentences (for some preliminary studies, see Collins (2024a, 2024b)). In some ways, ChatGPT's language use very closely resembles that of humans (Cai et al., 2024; Ortega-Martín et al., 2023; Wang et al., 2023); however, there are many ways in which ChatGPT and other LLMs like it fail to linguistically perform on the same level as humans (Basmov, Goldberg, & Tsarfaty, 2024; Borji, 2023; Jang & Lukasiewicz, 2023; Shen, Chen, Backes, & Zhang, 2023; Zhong et al., 2023; Zuccon & Koopman, 2023, a.o.). These findings suggest that, while ChatGPT and other LLMs are able to generate human-like linguistic utterances, they do not have the same mechanism of generating internal hierarchical syntactic structures that human language is endowed with (Contreras Kallens et al., 2023; Hale & Stanojević, 2024; Linzen & Baroni, 2021; Manova, 2024b; H. Zhou et al., 2023, a.o.).

Emojis are pictorial elements that exist in the digital keyboard layout of most smartphones and computers. They are easily integrated with text and, as such, appear often in written language in computer-mediated communication (CMC). Emojis often appear in written utterances as morphosyntactic constituents Storment (2024), a phenomenon known as pro-text emojis. There is strong evidence to suggest that pro-text emojis are fully integrated into linguistic systems in terms of their morphosyntax combinatorics (Homann, Brady R. T., Sara, & Fernandes, 2022; Stamatov, 2017; Storment, 2024), processing (Paggio & Tse, 2022; Scheffler, Brandt, Fuente, & Nenchev, 2022; Weissman, Engelen, Baas, & Cohn, 2023), and their mechanisms of semantic interpretation (Cohn, Roijackers, Schaap, & Engelen, 2018; Tieu et al., 2025). Emojis are unique as an incredibly widespread, conventionalized set of symbols and pictograms that can be readily incorporated into written natural language. In this paper I refer specifically to pictorial symbols accessible from a digital keyboard layout and associated with a specific Unicode point as "emojis", so this definition excludes traditional "emoticons" composed of preexisting text symbols such as :) and :(. These elements offer a very meaningful glimpse at the way that human language interfaces with visual information.

In this paper I present novel data from GPT-4o (OpenAI et al., 2024) handling sentences containing pro-text emojis in English and in Spanish. I show that, while GPT-4o is able to accurately identify what many – but crucially not all – emojis mean, it is completely unable to predict with any accuracy where pro-text emojis can acceptably appear in written utterances. This is the first study of its kind, as research on the morphosyntactic combinatorics of pro-text emojis is incredibly limited, and research on the metalinguistic awareness of LLMs giving grammaticality judgments is still relatively new at the time of authorship of this paper. Given these considerations, this article outlines a preliminary case study of how one LLM (GPT-4o) handles a relatively small sample of data from two languages (both of which have abundant data online) and how it shows notable differences from the way in which it conveys metalinguistic acceptability judgments on sentences containing pro-text emojis when compared to how it rates sentences which lack emojis. Future work on this topic should address similar questions using a significantly larger sample size of data from significantly more languages with tests run on multiple models, not just GPT-4o.

Human cognition allows for the extraction of meaning from visual stimuli, and this seems to be strongly connected to the human language faculty when we look at the existence of language phenomena such as iconicity, partial iconicity, pictorial symbols in orthography, gesture, and, of course, the existence of signed languages. ChatGPT is crucially missing that aspect of the human capacity for language, and it seems to have no functional alternative, even for visual elements that appear in text. This gap – along with the gaps in syntactic awareness, which has been well-documented – spells doom for LLMs when it comes to the treatment of visual elements as syntactic constituents. This is a gap that must be filled if generative AI models are to truly match human language performance.

2 Pro-text emojis

Pro-text emojis (borrowing terminology from the semantics of gestures in spoken language (Schlenker, 2019)) are emojis that appear inside of an utterance as morphosyntactic constituents. See the following examples from Storment (2024), ultimately from Twitter/X.

(1) a. I need to 🛌 before I see the end of this game or I'll be 😢 I missed it

b. 🏠 is where the ❤️ is

c. Some 🌈 people were discriminated against at protest grounds

This phenomenon is not restricted to English. See the following examples from Spanish, also taken from Twitter/X.

(2) a.

Mejor	me	voy	a	😴...
better	me	go.1sg	to	😴...

'I'd better go to sleep...'

b.

Y	que	se	vayan	a	tu	país	los	que	odian	el 🇵🇾
and	that	REFL	go.3pl.sbj	to	your	country	those	that	hate	the 🇵🇾

'And that those who hate Paraguay go to your country'

c.

Mi	solidaridad	con	la	gente	🏳️‍⚧️	y	no-binaria siempre!
my	solidarity	with	the	people	🏳️‍⚧️	and	non-binary always!

'My solidary with trans and non-binary people always!'

Some work on the semantics of pro-text emojis suggest that they replace written words in an utterance (Tieu et al., 2025), but this conclusion is puzzling as pro-text emojis may appear as elements smaller than a completely formed word in a given derivation in both English and in Spanish. Examples again from Twitter/X.

(3) a. He ❤️s to 📚

b. My therapist 👻ed me so I took selfies in the parking lot

c. likeeee the secondhand embarrassment is ☠️ing me

d. A couple of 🚬s smoking 🚬s

(4) a.

Las	🍓s	me	encantaaan
the.fpl	🍓pl	me	love.3pl

'I looove strawberries'

b. Un ☕cito para reconfortar
a.msg ☕DIM for refresh
'A little coffee to refresh you'

It is not the case, though, that pro-text emojis may freely replace any morpheme. Storment (2024) systematically demonstrates that there are licit and illicit morphosyntactic positions in which pro-text emojis are licensed, and language users have clear grammaticality judgments about what these positions are.

(5) a. *I like 🌸al perfumes (int: floral)

b. *Professor Rambow is a 💻ational linguist (int: computational)

c. *🤨ity killed the cat (int: curiosity)

d. *I need to 💪en my core (int: strengthen)

The examples in (5) are ungrammatical because such forms are unattested online and because English-speaking emoji users consistently judge these forms to be ungrammatical.

There are also restrictions on pro-text emoji placement in other languages, though the restrictions vary from language to language. Take the example of Spanish.

(6) a. Yo te ❤️(*-o) mucho
I you ❤️(*-1SG.PRES) much
'I love you very much'

b. Tú me ❤️(*-s) mucho
you me ❤️(*-2SG.PRES) much
'You love me very much'

While verbal agreement and tense suffixes are possible with pro-text emojis in English (2a-c), they are generally not possible in Spanish. This shows that the combinatorics of pro-text emojis with other morphemes are sensitive to the morphosyntactic structure of the language in which they are embedded. See Storment (2024) for a detailed analysis of this difference between English and Spanish.

These data clearly show two very important facts. First, pro-text emojis do not replace orthographic words. They appear as units smaller than the word-level, yet it is not the case that they can freely stand in for any morpheme. As such, they take part in morphosyntactic operations and must be obedient to the grammatical constraints of the language in which they appear, which is the second important fact.

3 GPT-4o tests

I used OpenAI's ChatGPT interface to perform a series of linguistic experiments with GPT-4o on sentences containing pro-text emojis. I performed these experiments on two languages: English and Spanish. I tested GPT-4o's knowledge on emoji recognition in isolation, emoji recognition embedded in sentences, and metalinguistic acceptability judgments on sentences containing pro-text emojis. I detail the findings of these short experiments here.

3.1 Emoji recognition

GPT-4o is, for the most part, exceptionally good at recognizing emojis. This is consistent with prior research (Y. Zhou, Lu, Gao, Mei, & Ai, 2024). It can identify what an emoji depicts, how it is used, and even what an emoji's Unicode code sequence is. I asked GPT-4o to identify what a given emoji depicts, what its keywords are, and what its Unicode entry is for a random sample of 40 emojis, five from each category of emoji: Smileys and People, Animals & Nature, Food & Drink, Activity, Travel & Places, Objects, Symbols, and Flags. I ran this prompt (shown in (7), followed by a numbered list of forty emojis) three times.

(7) *Emoji recognition prompt*
Identify the meaning, keywords, and Unicode code associated with the following emojis.

GPT-4o was able to identify all information correctly for all emojis in all three tests. In other words, it performed perfectly.

However, during some preliminary testing in which I gave GPT-4o the prompt in (8), some of the results were quite different.

(8) *Preliminary testing for emoji recognition*
What is the metadata for the emoji _____? I mean the Unicode entry and keywords that are associated with each emoji.

I did this for thirteen random emojis, and it did not perform perfectly. It failed to identify three emojis: 🦷 (tooth), 👂 (ear), and 🦵 (leg). In addition to misidentifying the 🦷 emoji, the Unicode code it gave here is for the safety pin (🧷) emoji. At the time of writing this, there currently is no pigeon emoji. The Unicode code for the 👂 emoji given here was correct, despite misidentifying the meaning. GPT 4-o identified the giraffe emoji as a "monkey face emoji", and, while there is a monkey face emoji that exists, the Unicode code given was the one for the giraffe (🦒) emoji. GPT-4o is mostly able to correctly identify emojis, but it still makes mistakes. The mistakes seem inconsistent and unpredictable, though I was able to force it to misidentify these three emojis several more times outside of the context of this formal experiment. Interestingly, the three emojis that I found GPT-4o to struggle with all depict a part of the human body.

3.2 Translation

For the next task, I had GPT-4o "translate" sentences containing pro-text emojis into sentences that only contain standard orthographic words. Using both grammatical and ungrammatical forms, I gave GPT-4o the following prompt along with 40 sentences containing pro-text emojis.

(9) *Translation prompt*
Paraphrase each of the following sentences without using emojis.

In the prompt, I did not indicate which sentences were grammatical and ungrammatical. GPT-4o executed this task perfectly, and accurately conveyed the meaning – or indented meaning – of each sentence. This study demonstrates that GPT-4o has a solid grasp on the semantic content of emojis and their conventions of use, though there is no indication here that it has any notion of syntax or at least morphological combinatorics. I confirm this in the following experiment.

3.3 Acceptability judgments

I then had GPT-4o rate grammatical and ungrammatical utterances for their acceptability. I did this for 50 English sentences and 35 Spanish sentences containing pro-text emojis. I ran each prompt three times. I found that GPT-4o is inaccurate and inconsistent when it comes to giving grammaticality judgments of sentences containing pro-text emojis, and these judgments do not match those from human native speakers.

I gave GPT-4o the following prompt(s). I had to specify "informal English/Spanish" because otherwise it judged almost every sentence to be unacceptable.

(10) *Acceptability judgment prompt*
For each sentence, tell me if it is acceptable in informal English/Spanish or not. Do not give any explanations.

One reviewer points out that "acceptable" in the context of this prompt is a very general term. Acceptable could mean anything from logically acceptable to politically acceptable. While this is true, the fact that other studies (e.g. Collins, 2024a, 2024b) as well as preliminary experimentation for this project use the term acceptability to refer to human-like intuitions about the grammaticality of certain linguistic forms, and that this is crucially something that the models seem to grasp in these experiments as they give accurate answers, is at least somewhat indicative that the models understand what acceptability refers to in this context.

Preliminary data from Collins (2024a, 2024b) show that GPT-4o is quite good at giving native-speaker-like grammaticality judgments for sentences that do not contain pro-text emojis, and my data here show that GPT-4o's judgments are quite inaccurate for emoji sentences. I assigned each sentence in each language a match score, which was determined by how many times GPT-4o accurately gauged the acceptability of each sentence across all three trials (i.e., a match score of 66.6667 indicates that it gave an accurate response 2/3 times). Here I show some examples of grammatical and ungrammatical forms in both English and Spanish that were used in this experiment, as well as the GPT-4o match score for each of these sentences in parentheses.

(11) a. I ❤️ you (100)
b. He is the 🏳️‍🌈est person ever (100)
c. I love 🥭es (0)
d. Wow, 👊 must really love yourself (100)

(12) a. *My son is learning ➕ition in school (66.6667)
b. *Mr. Kamano is 🇯🇵ese (66.6667)
c. *Wow, you must really love 👊self (33.3333)

(13) a. Te ❤️ mucho (100)
b. Aquí están los 🐶itos (100)
c. Las 🍎s son rojas (66.6667)

(14) a. *Tú me ❤️s mucho (0)
b. *Las 🍎as son rojas (100)
c. *Yo te ❤️é antes (33.3333)

Acceptability (i.e., grammatical sentences) for these input sentences was determined by attestation online, as well as being judged as grammatical by native speakers of each language in medium-scale judgement tasks with 10-15 participants. In these tasks, native speakers of each language were simply asked to rate sentences as either grammatical or ungrammatical. Interestingly, Spanish speakers felt more strongly about the ungrammaticality of the ungrammatical examples than did English speakers, an observation mirrored in Storment (2024). These data are supported by the lack of attestation for ungrammatical forms. Sentences shown here marked as grammatical were unanimously ranked as such, and the same is true for the ungrammatical utterances. In other words, the sentences in (11) and (13) are marked as acceptable because they are forms that native speakers produce and they are forms that native speakers say are grammatical. Ungrammatical examples such as those in (12) and (14) are marked as such because they are unattested online and because native speakers of English and Spanish report that these sentences are not grammatical. Judgements from human language users are clear on utterances containing pro-text emojis. Judgments from both humans and LLMs on sentences which do not contain pro-text emojis are also clear. This experiment demonstrates that LLM judgments on sentences containing pro-text emojis are what is least clear.

The average match score for all English sentences is 63.3332, and the average match score for all Spanish sentences is 69.5238. GPT-4o performed slightly better with the

Spanish examples than it did with the English examples. Interestingly, this is parallel with the observation in Storment (2024) that native Spanish speakers have more robust judgments on sentences containing pro-text emojis in Spanish than do English speakers in English. More research and experimentation is necessary to confirm if there is actual correlation there, or if it is just coincidence.

Overall, GPT-4o gave inaccurate ratings for these sentences. It especially struggled to label the ungrammatical examples as such, though it also inaccurately labeled many grammatical sentences. In the following section I discuss the theoretical implications of this both for human language and for generative AI systems.

4 Theoretical implications

Emoji combinatorics – that is, where an emoji may appear within a given utterance – reveal a hierarchical internal syntax that these emojis must abide by. There is some internal hierarchical structure in each language that licenses these visual elements in some locations, but not others. The distribution of these elements forces us to consider the syntactic structures of natural language.

LLMs, however, do not operate over hierarchically-ordered syntactic constituents in the exact same way that human language does (Manova, 2024a, 2024b), and may not do so at all with pictorial constituents. This raises two important questions. First, what parameters does the LLM use to determine the grammaticality of these utterances containing pro-text emojis? Second, how does it determine the grammaticality of any utterance? The idea is that humans use the same metrics to determine the grammaticality of the emoji utterances than they would any other, but it is not clear that that is the case for LLMs, especially given how relatively rare these kinds of data are. LLMs like GPT-4o determine the grammaticality of utterances by generalizing over the data in their training sets. If a given form frequently appears in the system's training data, it is more likely to accept such forms as grammatical. Conversely, if a given form is poorly attested or completely unattested in the training data, it is less likely to accept those forms.

Frequency and attestation is one of the metrics used in Storment (2024) to determine the acceptability of a given pro-text emoji form. In that paper, grammatical forms are well-attested online (in addition to being judged grammatical by native speakers), and ungrammatical forms are either poorly attested or not attested (also in addition to being judged ungrammatical by native speakers). As shown in the numerous studies cited in the introduction of this paper, the training data works particularly well in many cases to cause the LLM to come to the correct conclusions on grammaticality, though it is not perfect. In fact, the metrics of attestation and frequency are quite accurate in Storment (2024) for determining the grammaticality of emoji sentences, so one might expect a greater degree of accuracy from the LLMs in this way. This forces us to wonder why GPT-4o struggles particularly with these data containing pro-text emojis, assuming it contains such utterances in its training set. If it does not, then this is an obvious way in which the model can improve, though I assume it does have some exposure to this kind of data because it can interpret the sentences successfully and because the model is trained on 175 billion parameters. It could, of course, also be the case that these forms are not in the training data, or at least not enough for the model to meaningfully abstract over the data. This would explain why GPT-4o struggles with some grammatical (i.e., attested) examples as well. This would make sense given that, of all the written utterances online, utterances containing emojis are relatively infrequent. Furthermore, pro-text emojis, despite being used by many language users, are perhaps the most infrequently-occurring use of emojis when compared to things like pro-text and co-text emojis (Storment, 2024; Tieu et al., 2025). It could very well be the case that GPT-4o simply isn't very familiar with these forms, but I consider other options here as to maximally inform a theory of LLM language use and human language cognition.

Generative AI systems do not have the capacity for processing iconic and pictorial elements, at least not directly. Obviously, they do not have the means of directly perceiving and processing visual and auditory stimuli in the way that that humans do. Whatever iconicity is (Davidson, 2023), and however the human brain processes that information, LLMs lack that same resource. One has to wonder if the improvement of an LLM's ability to process visual information would correlate with an improvement in the model to accurately gauge the acceptability of sentences containing pictorial constituents.

Human language readily integrates iconic and pictorial elements, be it depictive sounds (Wiese, 1996), gestures (Goldin-Meadow & Brentari, 2017; Schlenker, 2019), emojis (Grosz, Kaiser, & Pierini, 2021; Storment, 2024), or a more traditional notion of "pictures". Storment (2024) shows that the way in which these elements are incorporated into language is indicative of some underlying internal structure (this idea is expressed elsewhere, such as Wiese (1996)). The ability to incorporate iconic visual elements into language perhaps relies on the ability of these elements to be embedded in a syntax. In other words, the distribution and combinatorics of pictorial morphosyntactic constituents in language reveals certain considerations on the morphosyntactic structures of the languages that these elements appear in. Given that the hierarchical internal structure that LLMs assign to utterances is not done in the same way that humans do it (Contreras Kallens et al., 2023; Hale & Stanojević, 2024; Linzen & Baroni, 2021; Manova, 2024a, 2024b; Y. Zhou et al., 2024, a.o.), and given that the distribution of pictorial elements in linguistic utterances is both dependent on and indicative of hierarchical internal structure, it stands to reason that GPT-4o struggles with giving human-like acceptability judgments for sentences containing pro-text emojis when it normally does not struggle with such a thing is because it is unable to make human-like generalizations about the internal syntactic structure of these pictorial elements specifically. The potential lack of attestation of these forms in GPT-4o's training data – along with the lack of a human-like syntax – could explain why the model is unable to generalize over these data containing pro-text emojis.

5 Conclusion

The use of emojis and other pictorial elements is an informatic tool, a tool that forces us to consider the morphosyntactic constraints of a given language. We can use this tool to see where the limitations of LLMs' comprehension lie.

In this paper I show that the LLM GPT-4o cannot make accurate generalizations about the grammaticality of visual elements that are embedded inside utterances in English and Spanish, even though it, for the most part, can very accurately interpret the semantic content of emojis and the utterances in which they appear.

This paper is meant to lay the foundations for future research on this subject. The data and theoretical discussion presented here are still very preliminary, and they introduce many interesting questions concerning iconicity in syntax, human cognition, and the improvement of generative AI.

References

Basmov, V., Goldberg, Y., & Tsarfaty, R. (2024, April). *Simple Linguistic Inferences of Large Language Models (LLMs): Blind Spots and Blinds.* arXiv. Retrieved 2025-06-03, from `http://arxiv.org/abs/2305.14785` (arXiv:2305.14785 [cs]) doi: 10.48550/arXiv.2305.14785

Borji, A. (2023, April). *A Categorical Archive of ChatGPT Failures.* arXiv. Retrieved 2025-06-03, from `http://arxiv.org/abs/2302.03494` (arXiv:2302.03494 [cs]) doi: 10.48550/arXiv.2302.03494

Cai, Z. G., Duan, X., Haslett, D. A., Wang, S., & Pickering, M. J. (2024, March). *Do large language models resemble humans in language use?* arXiv. Retrieved 2025-06-03, from `http://arxiv.org/abs/2303.08014` (arXiv:2303.08014 [cs]) doi: 10.48550/arXiv.2303.08014

Cohn, N., Engelen, J., & Schilperoord, J. (2019, August). The grammar of emoji? Constraints on communicative pictorial sequencing. *Cognitive Research: Principles and Implications*, *4*(1), 33. Retrieved 2025-06-03, from `https://doi.org/10.1186/s41235-019-0177-0` doi: 10.1186/s41235-019-0177-0

Cohn, N., Roijackers, T., Schaap, R., & Engelen, J. (2018). Are emoji a poor substitute for words? {Sentence} processing with emoji substitutions. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 40). Retrieved from `https://escholarship.org/uc/item/33d802h8`

Collins, C. (2024a, October). *Ordinary Working Grammarian: Acceptability Judgments in ChatGPT.* Retrieved 2025-06-03, from `https://ordinaryworkinggrammarian.blogspot.com/2024/10/acceptability-judgments-in-chatgpt.html`

Collins, C. (2024b, October). *Ordinary Working Grammarian: Acceptability Judgments in ChatGPT (Study 2).* Retrieved 2025-06-03, from `https://ordinaryworkinggrammarian.blogspot.com/2024/10/acceptability-judgments-in-chatgpt_20.html`

Contreras Kallens, P., Kristensen-McLachlan, R. D., & Christiansen, M. H. (2023). Large Language Models Demonstrate the Potential of Statistical Learning in Language. *Cognitive Science*, *47*(3), e13256. Retrieved 2025-06-03, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13256 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.13256) doi: 10.1111/cogs.13256

Davidson, K. (2023). Semiotic distinctions in compositional semantics. *Proceedings of the 58th Meeting of the Chicago Linguistic Society*. Retrieved from https://projects.iq.harvard.edu/files/meaningandmodality/files/davidson2023-clsproceedings-semiotic.pdf

Goldin-Meadow, S., & Brentari, D. (2017, January). Gesture and language: Distinct subsystem of an integrated whole. *Behavioral and Brain Sciences*, *40*, e74. Retrieved 2025-06-03, from https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/gesture-and-language-distinct-subsystem-of-an-integrated-whole/66110F96CE4668BD5FD0918BE5E36258 doi: 10.1017/S0140525X1600039X

Grosz, P., Kaiser, E., & Pierini, F. (2021, September). Discourse anaphoricity and first-person indexicality in emoji resolution. *Proceedings of Sinn und Bedeutung*, *25*, 340–357. Retrieved 2025-06-16, from https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/941 doi: 10.18148/sub/2021.v25i0.941

Hale, J. T., & Stanojević, M. (2024, November). Do LLMs learn a true syntactic universal? In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17106–17119). Miami, Florida, USA: Association for Computational Linguistics. Retrieved 2025-06-03, from https://aclanthology.org/2024.emnlp-main.950/ doi: 10.18653/v1/2024.emnlp-main.950

Homann, L. A., Brady R. T., R., Sara, A., & Fernandes, M. A. (2022, April). Are emojis processed visuo-spatially or verbally? Evidence for dual codes. *Visual Cognition*, *30*(4), 267–279. Retrieved 2025-06-03, from https://doi.org/10.1080/13506285.2022.2050871 doi: 10.1080/13506285.2022.2050871

Jang, M. E., & Lukasiewicz, T. (2023, November). *Consistency Analysis of ChatGPT.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2303.06273 (arXiv:2303.06273 [cs]) doi: 10.48550/arXiv.2303.06273

Kennedy, M. (2025, May). Evidence of Generative Syntax in LLMs.. Retrieved 2025-06-03, from https://openreview.net/forum?id=v2w8LUpt0a

Linzen, T., & Baroni, M. (2021, January). Syntactic Structure from Deep Learning. *Annual Review of Linguistics*, *7*(Volume 7, 2021), 195–212. Retrieved 2025-06-03, from https://www.annualreviews.org/content/journals/10.1146/annurev-linguistics-032020-051035 (Publisher: Annual Reviews) doi: 10.1146/annurev-linguistics-032020-051035

Manova, S. (2024a, June). *ChatGPT and linguistic theory, with a focus on morphology.* LingBuzz. Retrieved 2025-06-03, from https://ling.auf.net/lingbuzz/008600 (LingBuzz Published In: Submitted for inclusion in José-Luis Mendívil-Giró (ed.), Artificial Knowledge of Language. A Linguists' Perspective on its Nature, Origins and Use. Wilmington, DE: Vernon Press.)

Manova, S. (2024b, November). *Machine learning versus human learning: Basic units and form-meaning mapping.* LingBuzz. Retrieved 2025-06-03, from `https://lingbuzz.net/lingbuzz/008548` (LingBuzz Published In: Submitted for inclusion in Vsevolod Kapatsinski and Gašper Beguš (eds.), Implications of Neural Networks and Other Learning Models for Linguistic Theory. Special issue of Linguistics Vanguard.)

OpenAI, Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., … Malkov, Y. (2024). *GPT-4o System Card.* arXiv. Retrieved 2025-06-20, from `https://arxiv.org/abs/2410.21276` (Version Number: 1) doi: 10.48550/ARXIV.2410.21276

Ortega-Martín, M., García-Sierra, O., Ardoiz, A., Álvarez, J., Armenteros, J. C., & Alonso, A. (2023, February). *Linguistic ambiguity analysis in ChatGPT.* arXiv. Retrieved 2025-06-03, from `http://arxiv.org/abs/2302.06426` (arXiv:2302.06426 [cs]) doi: 10.48550/arXiv.2302.06426

Paggio, P., & Tse, A. P. P. (2022). Are Emoji Processed Like Words? An Eye-Tracking Study. *Cognitive Science*, *46*(2), e13099. Retrieved 2025-06-03, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13099` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.13099) doi: 10.1111/cogs.13099

Pierini, F. (2021, September). Emojis and gestures: a new typology. *Proceedings of Sinn und Bedeutung*, *25*, 720–732. Retrieved 2025-06-16, from `https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/963` doi: 10.18148/sub/2021.v25i0.963

Scheffler, T., Brandt, L., Fuente, M. d. l., & Nenchev, I. (2022, February). The processing of emoji-word substitutions: A self-paced-reading study. *Computers in Human Behavior*, *127*, 107076. Retrieved 2025-06-03, from `https://www.sciencedirect.com/science/article/pii/S074756322100399X` doi: 10.1016/j.chb.2021.107076

Schlenker, P. (2019, May). Gestural semantics. *Natural Language & Linguistic Theory*, *37*(2), 735–784. Retrieved 2025-06-03, from `https://doi.org/10.1007/s11049-018-9414-3` doi: 10.1007/s11049-018-9414-3

Shen, X., Chen, Z., Backes, M., & Zhang, Y. (2023, October). *In ChatGPT We Trust? Measuring and Characterizing the Reliability of ChatGPT.* arXiv. Retrieved 2025-06-03, from `http://arxiv.org/abs/2304.08979` (arXiv:2304.08979 [cs]) doi: 10.48550/arXiv.2304.08979

Stamatov, E. G. (2017). *Do emoji use a grammar? Emergent structure in non-verbal digital communication* (Master's thesis, Tilburg University). Retrieved from `https://arno.uvt.nl/show.cgi?fid=142877`

Storment, J. D. (2024, May). Going lexicon? The linguistic status of pro-text emojis. *Glossa: a journal of general linguistics*, *9*(1). Retrieved 2025-06-03, from `https://www.glossa-journal.org/article/id/10449/` (Number: 1 Publisher: Open Library of Humanities) doi: 10.16995/glossa.10449

Tieu, L., Qiu, J. L., Puvipalan, V., & Pasternak, R. (2025, April). Experimental evidence for a semantic typology of emoji: Inferences of co-, pro-, and post-text emoji. *Quarterly Journal of Experimental Psychology*, *78*(4), 808–826. Retrieved 2025-06-03, from https://doi.org/10.1177/17470218241255786 (Publisher: SAGE Publications) doi: 10.1177/17470218241255786

Wang, J., Hu, X., Hou, W., Chen, H., Zheng, R., Wang, Y., … Xie, X. (2023, August). *On the Robustness of ChatGPT: An Adversarial and Out-of-distribution Perspective.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2302.12095 (arXiv:2302.12095 [cs]) doi: 10.48550/arXiv.2302.12095

Weissman, B., Engelen, J., Baas, E., & Cohn, N. (2023). The Lexicon of Emoji? Conventionality Modulates Processing of Emoji. *Cognitive Science*, *47*(4), e13275. Retrieved 2025-06-03, from https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.13275 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.13275) doi: 10.1111/cogs.13275

Wiese, R. (1996, January). Phrasal compounds and the theory of word syntax. *Linguistic Inquiry*, *27*, 183–193.

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023, March). *Can ChatGPT Understand Too? A Comparative Study on ChatGPT and Fine-tuned BERT.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2302.10198 (arXiv:2302.10198 [cs]) doi: 10.48550/arXiv.2302.10198

Zhou, H., Hou, Y., Li, Z., Wang, X., Wang, Z., Duan, X., & Zhang, M. (2023, November). *How Well Do Large Language Models Understand Syntax? An Evaluation by Asking Natural Language Questions.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2311.08287 (arXiv:2311.08287 [cs]) doi: 10.48550/arXiv.2311.08287

Zhou, Y., Lu, X., Gao, G., Mei, Q., & Ai, W. (2024, April). *Emoji Promotes Developer Participation and Issue Resolution on GitHub.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2308.16360 (arXiv:2308.16360 [cs]) doi: 10.48550/arXiv.2308.16360

Zuccon, G., & Koopman, B. (2023, February). *Dr ChatGPT, tell me what I want to hear: How prompt knowledge impacts health answer correctness.* arXiv. Retrieved 2025-06-03, from http://arxiv.org/abs/2302.13793 (arXiv:2302.13793 [cs]) doi: 10.48550/arXiv.2302.13793

Correspondence

John David Storment

Stony Brook University
Department of Linguistics
Stony Brook, NY, USA
john.david.storment@stonybrook.edu