Journal of Pan-Pacific Association of Applied Linguistics, 27(2), 1–12.
https://doi.org/10.25256/PAAL.27.2.1

Does Google Bard Understand 'Itself'?

Sang-Gu Kang[*]

Kang, S.-G. (2023). Does Google Bard understand 'itself'? *Journal of Pan-Pacific Association of Applied Linguistics, 27*(2), 1–12.

Generative AIs such as Google Bard are known to be equipped with techniques and grammatical principles of human language based on a large corpus of text and code that allow them to generate natural-sounding language, and also identify and correct grammatical errors in human-written texts. Still, they are not perfect language generators, and this paper reports some interesting errors produced by Google Bard regarding interpretation of English plain and reflexive pronouns. Although Google Bard is clearly well aware of how to locate the antecedents for English pronouns, it consistently produces seemingly random interpretation errors when the prompt asks it to create a short story based on test sentences containing a plain or a reflexive pronoun. Since it is quite evident that Google Bard does not have an inborn system of grammatical principles including the binding principles, which are used to account for how human beings deal with pronoun interpretation, other approaches to processing human language need to be considered to explain Google Bard's errors and the underlying mechanisms leading up to the errors. Thus, a processing account based on the Emergentist approach (O'Grady, 2005) and the Emergentist Reflexivity Approach model (Sperlich, 2020) is adopted, and the two separate levels of processing, that is, sentence processing and pragmatic processing, are introduced to account for Google Bard's errors and the human language interpretation mechanism that might be underlying. These findings can not only shed light on how generative AIs might be analyzing human language but also contribute to our better understanding of human language and related theories.

Keywords: generative AIs, English pronoun interpretation, binding principles, Emergentism, sentence processing and pragmatic processing

1 Introduction

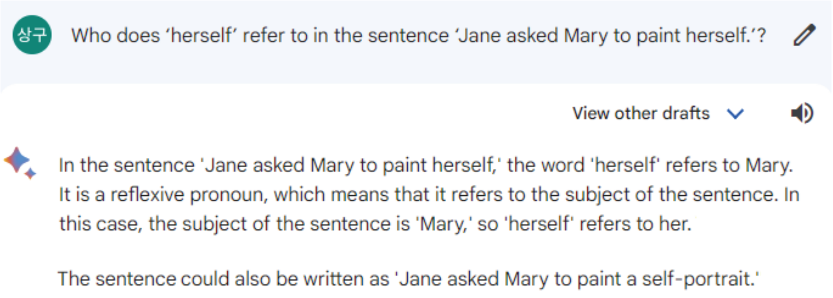
With the advent of generative AIs such as Chat GPT and Google Bard in recent years, our expectations about the potential to revolutionize diverse aspects of our lives have been raised. Generative AIs are known to be capable of a wide range of tasks including but not limited to generating creative contents, translating languages, providing informative answers to diverse questions, developing new products and services, and generating new ideas for scientific

[*] **Sang-Gu Kang**, Assistant Professor, Department of English Language and Literature, Gangneung-Wonju National University

research, all of which inevitably involves understanding and generating human language. Even some of the known potential risks associated with generative AIs including creating fake news or deepfakes that could be used to deceive or manipulate people cannot be created without the AIs understanding and generating human language. In addition to a number of techniques to generate grammatically correct texts, generative AIs are equipped with grammatical principles based on a large corpus of text and code that allow them to generate natural-sounding human language. However, there is a lack of literature dealing with the grammatical errors committed by AIs.[1] In this research, Google Bard's ability to understand a phenomenon of human language considered as one of the more complicated structures, interpreting pronouns, will be probed.

Understanding how pronouns are interpreted has been one of the key issues in the field of language acquisition as "[it] provides us with a probe into the basic mechanisms underlying linguistic computation, and … ways to differentiate between the contributions the various components of the language system make" (Reuland, 2011). One of the key observations regarding cross-linguistic interpretation of pronouns involves languages like English only allowing local antecedent for reflexive pronouns while East Asian languages like Chinese, Japanese, and Korean allowing both local and long-distance antecedents. Thus, in English sentence (1) the reflexive pronoun *herself* can only refer to the local antecedent Mary but not to the long-distance antecedent Jane, while the corresponding sentences in East Asian languages would allow their reflexive pronouns to refer to either Mary or Jane.

(1) Jane$_i$ asked Mary$_j$ to paint herself$_{*i/j}$.



Who does 'herself' refer to in the sentence 'Jane asked Mary to paint herself.'?

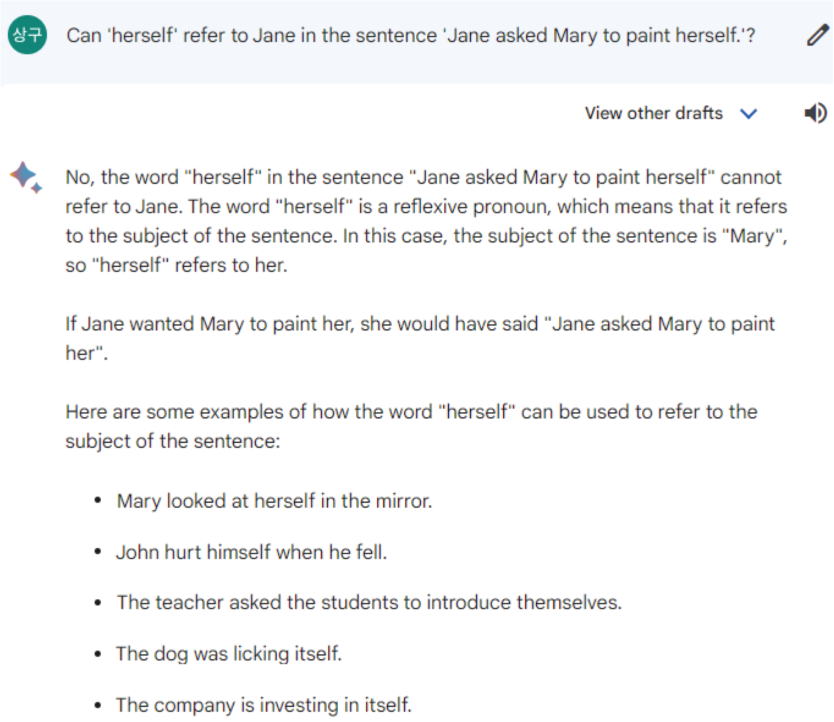
In the sentence 'Jane asked Mary to paint herself,' the word 'herself' refers to Mary. It is a reflexive pronoun, which means that it refers to the subject of the sentence. In this case, the subject of the sentence is 'Mary,' so 'herself' refers to her.

The sentence could also be written as 'Jane asked Mary to paint a self-portrait.'

Figure 1. Google Bard's response to the prompt "Who does 'herself' refer to in the sentence 'Jane asked Mary to paint herself.'?"

[1] Although there are studies on AIs as tools for improving grammar competence, there is, to the author's knowledge, no prior research on pronoun interpretation errors produced by AIs. However, there are studies such as Ortega-Martin et al. (2023) probing other linguistic challenges that generative AIs face, linguistic ambiguity.

Google Bard seems to be aware of the fact that English reflexives only allow local binding as evidenced by its responses to successive questions asking how an English reflexive pronoun *herself* should be interpreted. Figure 1 shows Google Bard's response to a prompt directly asking who the reflexive pronoun refers to in (1) above: it asserts that *herself* refers to Mary. A follow-up question asked whether *herself* can refer to Jane in Figure 2. It firmly denies the possibility answering that *herself* cannot refer to Jane but refers to Mary. It further goes on to say that if you want Mary to paint Jane, the sentence should be 'Jane asked Mary to paint her.'



Can 'herself' refer to Jane in the sentence 'Jane asked Mary to paint herself.'?

No, the word "herself" in the sentence "Jane asked Mary to paint herself" cannot refer to Jane. The word "herself" is a reflexive pronoun, which means that it refers to the subject of the sentence. In this case, the subject of the sentence is "Mary", so "herself" refers to her.

If Jane wanted Mary to paint her, she would have said "Jane asked Mary to paint her".

Here are some examples of how the word "herself" can be used to refer to the subject of the sentence:

- Mary looked at herself in the mirror.
- John hurt himself when he fell.
- The teacher asked the students to introduce themselves.
- The dog was licking itself.
- The company is investing in itself.

Figure 2. Google Bard's response to the prompt "Can 'herself' refer to Jane in the sentence 'Jane asked Mary to paint herself.'?"

Thus, it seems that Google Bard can understand and explain how English pronouns should be interpreted as suggested by Figures 1 and 2. Based on this fact, this paper investigates whether Google Bard shows the same degree of understanding for English pronouns in another type of task: asking it to create stories based on a given sentence that includes a pronoun. Then, the underlying mechanism for how Google Bard might actually be interpreting English pronouns is suggested.

2 Theoretical Background

In the generative tradition, Binding Principles A and B are two of the core principles of Binding Theory (Chomsky, 1981), which deals with syntactic constraints on the binding of anaphors and pronouns. Simply put, Binding Principle A states that an anaphor, which includes reflexive pronouns and reciprocal pronouns, requires a c-commanding antecedent in its binding domain, while Binding Principle B states that a pronoun cannot have a c-commanding antecedent in its binding domain.[2] Therefore, Binding Principle A explains how *herself* only refers to Mary in (1) (repeated from above), while Binding Principle B accounts for the fact that the plain pronoun *her* in (2) may refer to the long-distance antecedent Jane but not to the local antecedent Mary. However, the Binding Principles cannot universally apply to all languages as East Asian languages such as Chinese, Japanese, and Korean are well-known for allowing both the local and long-distance antecedents to bind their reflexive pronouns. Also related to the topic of this research, it should be noted that it does not make sense to say that generative AIs are encoded with the binding principles like human beings are argued to be (e.g., Chomsky, 1981).

(1) Jane$_i$ asked Mary$_j$ to paint herself$_{*i/j}$.
(2) Jane$_i$ asked Mary$_j$ to paint her$_{i/*j}$.

Other processing based approaches such as the Emergentist approach (O'Grady, 2005) and the Emergentist Reflexivity Approach model (Sperlich, 2020) that provide alternatives to the Chomskyan syntactic theories can account for how East Asian languages interpret reflexive pronouns, and ultimately embrace how pronouns are interpreted in English-type languages. These non-syntactic approaches posit two separate systems at work in processing anaphora: a sentence processor responsible for combining lexical items into phrases and sentences, and a pragmatic processor dealing with how context can influence interpretation.[3] The sentence processor aims to resolve dependencies arising from certain words (e.g., the verb *paint* will require two arguments: an agent and a theme) and to reduce the burden on working memory as quickly as possible. This type of processing is what Rao and McMahon (2022) suggested as a possibly more useful mechanism for machine learning compared to the ones based on traditional syntax. The sentence

[2] There is some debate about how to define the binding domain and how to account for certain apparent violations of the Binding Principles, which will not be discussed in detail as they are not included as the scope of this research.

[3] The specific and detailed differences between the non-syntactic approaches will not be discussed here, as they are not directly relevant to the topic of this research. The presence of the two separate processing systems that these approaches seem to agree on is of central importance. Refer to the references above for the exact mechanisms of the approaches and further discussion.

processor is at work in English-type languages and aims to immediately resolve the referent of an English reflexive pronoun, only allowing local binding. On the other hand, pragmatic processing is the default strategy in East Asian languages. Therefore, it will even allow the reflexive pronoun (in East Asian languages) to refer to an extra-sentential antecedent if it is prominent in the discourse in spite of the presence of a more local antecedent (O'Grady, 2013).

Then, English speakers will learn to mostly rely on the sentence processor and only allow local binding while East Asian language speakers will learn to rely on the pragmatic processor and allow both local and long-distance binding (O'Grady, 2013). Generative AIs like Google Bard also seem to be equipped with mechanisms similar to the two processors as the sentence processor will be necessary for the AIs to break down human language and analyze it, and the pragmatic processor can be used in establishing a clear meaning and message when analyzing the relationship between words. Thus, contemplating Google Bard's mechanism of interpreting pronouns, such as probing which of the two processors is more dominantly utilized, can provide an opportunity to better understand not only generative AIs but also human language itself.

3 Method

In order to probe how Google Bard interprets English pronouns, the researcher entered prompts asking it to create short stories based on a given sentence such as 'Jane asked Mary to paint her(self).' (above examples repeated as (3) and (4) here), in which two semantically plausible antecedents for the plain and reflexive pronouns were given. Thus, simple single sentence prompts such as 'Create a story in which Jane asked Mary to paint herself.' (Figure 3 in section 3 below) were typed into Google Bard. There were two types of test sentences: one containing a plain pronoun as in (3), and one containing a reflexive pronoun as in (4). Along with the feminine pronouns *her* and *herself*, masculine pronouns *him* and *himself* were also tested via prompts as in (5) and (6). In order to allow both the local and long-distance antecedents to be felicitous, two female characters Jane and Mary were introduced when feminine pronouns *her* and *herself* were used and two male characters Sam and Bill when masculine pronouns *him* and *himself* were used. Reflexive verbs *paint* and *shave* were selected to avoid violent contents in the stories created by Google Bard, and because they can create felicitous contexts involving two characters with the same sex.

(3) Jane asked Mary to paint her.
(4) Jane asked Mary to paint herself.
(5) Sam asked Bill to shave him.
(6) Sam asked Bill to shave himself.

Table 1. All 12 Test Sentences (Following 'Create a story in which …')

Main V	Reflexive V 'paint'	Reflexive V 'shave'
ask	Jane asked Mary to paint her.	Sam asked Bill to shave him.
	Jane asked Mary to paint herself.	Sam asked Bill to shave himself.
advise	Jane advised Mary to paint her.	Sam advised Bill to shave him.
	Jane advised Mary to paint herself.	Sam advised Bill to shave himself.
suggest	Jane suggested Mary to paint her.	Sam suggested Bill to shave him.
	Jane suggested Mary to paint herself.	Sam suggested Bill to shave himself.

Then, the main verb *ask* was used to create a natural context after embedding the clause containing the reflexive verb and the pronoun (e.g., 'Mary to paint her' in (3) above). Besides the four sentence above that used *ask* as the main verb, two more main verbs with similar meanings, *advise* and *suggest*, were used to create two extra sets of test sentences resulting in total 12 test sentences as in Table 1. These test sentences were embedded in an imperative matrix clause 'Create a story in which …' and given as prompts to Google Bard.

If Google Bard understands how the plain and reflexive pronouns should be interpreted, it would create a story in which Mary paints Jane when the plain pronoun *her* is given in test sentence (3) while a story in which Mary paints a self-portrait would be created when the reflexive *herself* is given in test sentence (4). In the same vein, it should create a story in which Bill shaves Sam's face when the plain pronoun *him* is given in test sentence (5) while a story in which Bill shaves his own face would be created when the reflexive *himself* is given in test sentence (6).

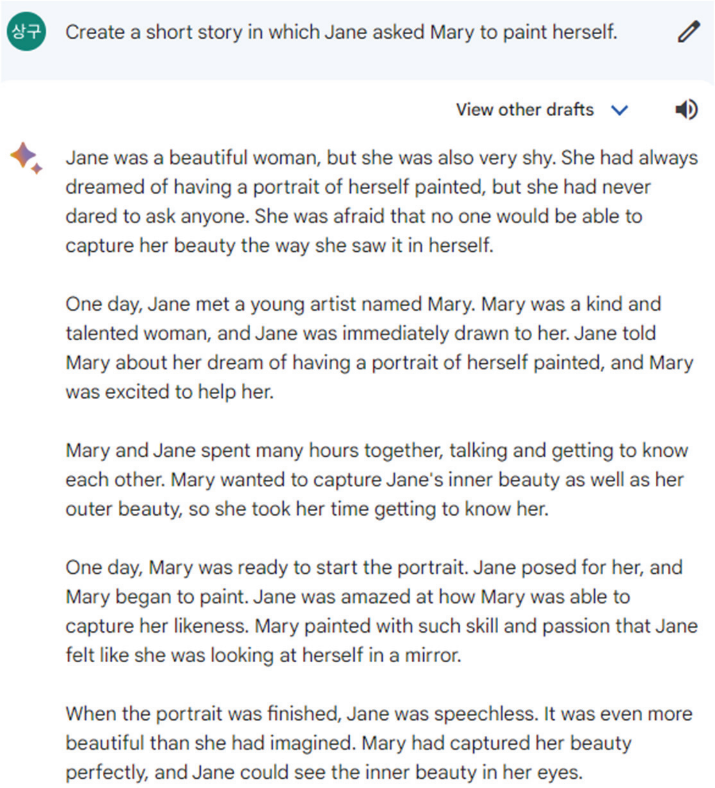
4 Results and Discussion

Although Google Bard created slightly different versions of the story for each test sentence every time the same prompt was given, the way it interpreted the (reflexive) pronoun was consistent within each test sentence. Therefore, a random example story for each test sentence was examined for this research.

Table 2. Pronoun Interpretation in Test Sentences Containing *paint*

Main V	Reflexive V 'paint'	Pronoun interpretation
ask	Jane asked Mary to paint her.	her → Jane
	Jane asked Mary to paint herself.	*herself → Jane
advise	Jane advised Mary to paint her.	*her → Mary
	Jane advised Mary to paint herself.	herself → Mary
suggest	Jane suggested Mary to paint her.	*her → Mary
	Jane suggested Mary to paint herself.	herself → Mary

Table 2 illustrates how the pronouns in the test sentences containing the reflexive verb *paint* were interpreted. Among the six stories based on the test sentences, three interpreted the pronouns correctly while the other three interpreted incorrectly; one incorrect interpretation involved the reflexive pronoun *herself* being treated as a plain pronoun and the other two involved the plain pronoun *her* being treated as a reflexive pronoun. Figure 3 is an example of Google Bard's creation that shows an incorrect interpretation of the reflexive pronoun *herself* in response to the prompt "Create a short story in which Jane asked Mary to paint herself." So although in Figure 1 above, Google Bard declared that *herself* should refer to Mary, in Figure 3 it created a story in which *herself* refers to Jane. The overall result suggests that Google Bard may not be equipped to conform to the grammar rules involving how pronouns should be interpreted because, on the surface, it looks as though the pronouns are randomly selecting their antecedents between the local and long-distance antecedents: not all plain pronouns are interpreted incorrectly and not all reflexive pronouns are interpreted correctly.



Create a short story in which Jane asked Mary to paint herself.

Jane was a beautiful woman, but she was also very shy. She had always dreamed of having a portrait of herself painted, but she had never dared to ask anyone. She was afraid that no one would be able to capture her beauty the way she saw it in herself.

One day, Jane met a young artist named Mary. Mary was a kind and talented woman, and Jane was immediately drawn to her. Jane told Mary about her dream of having a portrait of herself painted, and Mary was excited to help her.

Mary and Jane spent many hours together, talking and getting to know each other. Mary wanted to capture Jane's inner beauty as well as her outer beauty, so she took her time getting to know her.

One day, Mary was ready to start the portrait. Jane posed for her, and Mary began to paint. Jane was amazed at how Mary was able to capture her likeness. Mary painted with such skill and passion that Jane felt like she was looking at herself in a mirror.

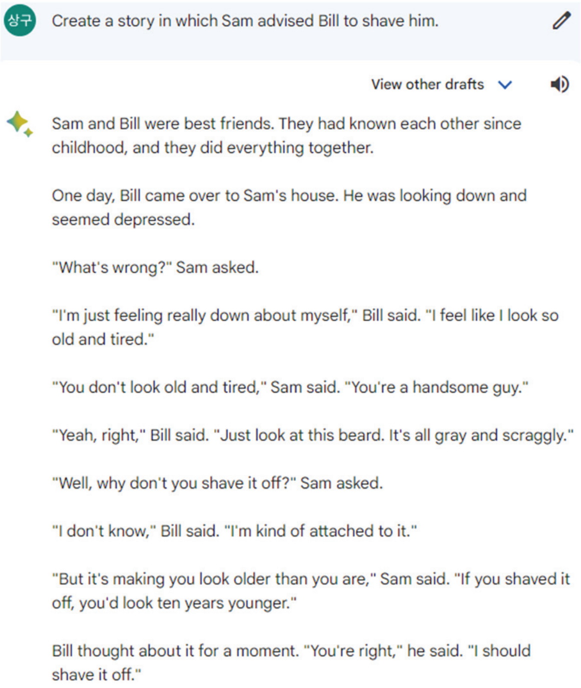
When the portrait was finished, Jane was speechless. It was even more beautiful than she had imagined. Mary had captured her beauty perfectly, and Jane could see the inner beauty in her eyes.

Figure 3. The beginning of Google Bard's response to the prompt "Create a short story in which Jane asked Mary to paint herself."

Next, Table 3 illustrates how the pronouns in the test sentences containing the reflexive verb *shave* were interpreted. Among the six stories based on the test sentences, four interpreted the pronouns correctly while the other two interpreted incorrectly; both incorrect interpretations involved the plain pronoun *him* being treated as a reflexive pronoun. Figure 4 is an example of Google Bard's creation that shows an incorrect interpretation of the plain pronoun *him* in response to the prompt "Create a short story in which Sam advised Bill to shave him." Just as with the pronoun interpretations in test sentences containing *paint*, it seems difficult to figure out the mechanism underlying Google Bard's pronoun interpretation.

Table 3. Pronoun Interpretation in Test Sentences Containing *shave*

Main V	Reflexive V 'paint'	Pronoun interpretation
ask	Sam asked Bill to shave him.	him → Sam
	Sam asked Bill to shave himself.	himself → Bill
advise	Sam advised Bill to shave him.	*him → Bill
	Sam advised Bill to shave himself.	himself → Bill
suggest	Sam suggested Bill to shave him.	*him → Bill
	Sam suggested Bill to shave himself.	himself → Bill



Create a story in which Sam advised Bill to shave him.

Sam and Bill were best friends. They had known each other since childhood, and they did everything together.

One day, Bill came over to Sam's house. He was looking down and seemed depressed.

"What's wrong?" Sam asked.

"I'm just feeling really down about myself," Bill said. "I feel like I look so old and tired."

"You don't look old and tired," Sam said. "You're a handsome guy."

"Yeah, right," Bill said. "Just look at this beard. It's all gray and scraggly."

"Well, why don't you shave it off?" Sam asked.

"I don't know," Bill said. "I'm kind of attached to it."

"But it's making you look older than you are," Sam said. "If you shaved it off, you'd look ten years younger."

Bill thought about it for a moment. "You're right," he said. "I should shave it off."

Figure 4. The beginning of Google Bard's response to the prompt "Create a short story in which Sam advised Bill to shave him."

There are a couple of observations that can be made based on the above results. First, although in principle the referents of the plain pronouns *her* and *him* could be someone not appearing in the test sentences, Google Bard always selected one of the two antecedents within the test sentences. Always selecting an in-sentence antecedent, which is what an English native speaker would normally do as well if no additional context besides the test sentence is given, seems to be the most plausible and logical choice as there is probably no reason for Google Bard to come up with a character not given in the prompt.

Second and more importantly, although Google Bard's pronoun interpretation may appear random, it seems possible to make a generalization regarding its choice of pronoun referent. It should be noted that when Google Bard interpreted the feminine pronouns, the referents for both the plain pronoun *her* and the reflexive pronoun *herself* are consistent within the pair of sentences sharing the same main verb: the referents are Jane when the main verb is *ask*, Mary when the main verbs are *advise* and *suggest* (Table 2). This trend continues when Google Bard interpreted the masculine pronouns *him* and *himself* as the referents are Bill when the main verbs are *advise* and *suggest*. The only exception to the above occurred when Google Bard interpreted both the masculine plain pronoun *him* and reflexive pronoun *himself* correctly when the main verb is *ask* (Table 3).

This second observation should be further analyzed in depth as it can shed light on the underlying mechanism for Google Bard's pronoun interpretation. Unlike human beings who are allegedly equipped with the inborn binding principles according to syntactic approaches to language, Google Bard obviously cannot have been created with such inborn principles that allow them to interpret pronouns appropriately. Instead, the so-called machine learning algorithms enable AIs to break down human communication into basic concepts that are subsequently reinterpreted to analyze the relationship between words to establish a clear message. Rao and McMahon (2022) note that the earlier versions of AI that were not as successful in generating human language as the recent versions were fed with rules based on Chomskyan theory. Language models based on various statistics and probabilistic techniques empowered machines to more effectively interpret and generate human language, and as a result, better interact with humans. Then, how can Google Bard interpret pronouns without a seemingly proper mechanism to do so?

In an attempt to account for how Google Bard might be interpreting pronouns, this paper turns to some linguistic approaches such as the aforementioned Emergentist approach (O'Grady, 2005) and the Emergentist Reflexivity Approach model (Sperlich, 2020), which posited the existence of a sentence processor and a pragmatic processor. In the pragmatic sense, Google Bard's (seemingly random) choice of referents in interpreting pronouns is actually consistent. In Table 2 above, the test sentences 'Jane asked Mary to paint *pronoun*.' pragmatically make more sense if Jane is Mary's client asking

Mary to paint Jane: Jane asking for Mary's self-portrait is not impossible but seems less likely. The choice between the two interpretations depends on the sentence processor, but if we assume Google Bard is not trained to value the sentence processor mechanism over the pragmatic processor mechanism, we can see why the two sentences end up being interpreted the same. On the other hand, the test sentences 'Jane advised Mary to paint *pronoun*.' pragmatically make more sense if Jane gives Mary an advice about Mary painting a self-portrait: Jane giving an advice to Mary to paint Jane is not impossible but less plausible. In the same vein, the test sentences 'Jane suggested Mary to paint *pronoun*.' pragmatically make more sense if Jane gives Mary a suggestion on Mary painting a self-portrait: Jane suggesting Mary to paint Jane seems less plausible. Thus, if Google Bard's pronoun interpretation is mainly based on the pragmatic processor, we can understand why it erred on interpreting pronouns on certain test sentences.

Interpretation of masculine pronouns in Table 3 above can be analyzed in a similar way. The test sentences 'Sam advised/suggested Bill to shave *pronoun*.' pragmatically make more sense if Sam gives Bill an advice/a suggestion about shaving Bill's face: Sam giving an advice/a suggestion to Bill about shaving Sam's face is not impossible but seem less plausible. However, the two possible interpretations for the test sentences 'Sam asked Bill to shave *pronoun*.' seem similarly plausible. When the plain pronoun *him* is used, Sam can be barber Bill's client telling Bill that Sam's face needs shaving. When the reflexive pronoun *himself* is used, Sam could be in a position to tell Bill that Bill's face needs shaving. It seems that Google Bard was able to correctly interpret the pronouns in the test sentences 'Sam asked Bill to shave *pronoun*.' since both situations are plausible.

Such claim that Google Bard values contextual plausibility of the sentence in interpreting pronouns can be buttressed by having it 'Create a story in which Bill advised Mary to paint him.' Google Bard correctly interpreted *him* as Bill because the pronoun *him* must refer to a masculine antecedent although Bill advising Mary to paint Bill is pragmatically less likely. Therefore, it created a specific plausible situation in which Bill persuaded Mary, who was hesitant to paint a portrait because she had never painted one before, to paint him. This contrasts with the test sentence 'Jane advised Mary to paint her.' in which the two antecedents were both feminine and Google Bard ended up incorrectly interpreting *her* as referring to Mary because it would create a better opportunity to generate a plausible situation.

Overall, Google Bard's (mis)interpretations of English plain pronouns and reflexive pronouns are not random. Assuming the Google Bard's interpretation mechanism resembles that of human's following linguistic approaches such as Emergentism and the Emergentist Reflexivity Approach model, Google Bard seems to heavily rely on the pragmatic processor to make the best sense out of the given input. This is probably why if one context is heavily favored over others, Google Bard will interpret the input to match the pragmatically more plausible scenario without recourse to the sentence processor. Only when multiple contexts are pragmatically plausible will the sentence processor contribute to pronoun interpretation.

5 Conclusion

A fortuitous finding that Google Bard could not create a story matching the correct interpretation of 'Jane asked Mary to paint herself.' subsequently led to discovering that it produced seemingly random errors in interpreting English plain and reflexive pronouns. Although Figures 1 and 2 demonstrated that it was fully aware of how English plain and reflexive pronouns should be interpreted, it repeatedly erred when creating short stories based on certain test sentences containing a plain or reflexive pronoun. This phenomenon could not be accounted for based on the Chomskyan binding principles since it is clear that Google Bard does not have an inborn system of grammatical principles including the binding principles. In order to provide an alternative account, the concept of sentence processor and pragmatic processor was borrowed from the Emergentist approach and Emergentist Reflexivity Approach model. If we assume Google Bard is dominated by the pragmatic processor, and the sentence processor is activated only if the pragmatic processor sees similar level of plausibility in multiple interpretations, the errors produced can be accounted for. Although this research is a report of a very small sample size from a generative AI, which can be upgraded anytime to correct its flaws, I hope this triggers more interesting findings on how generative AIs (mis)interpret human languages. It will inevitably expand our understanding of not only how generative AIs operate but also about human language and the theories involved.

References

Chomsky, N. (1981). *Lectures on government and binding*. Foris Publications.

Ortega-Martín, M., García-Sierra, Ó., Ardoiz, A., Álvarez, J., Armenteros, J. C., & Alonso, A. (2023). *Linguistic ambiguity analysis in ChatGPT.* arXiv. https://doi.org/10.48550/arXiv.2302.06426

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Lawrence Erlbaum.

O'Grady, W. (2013). Processing and language acquisition: Reflexive pronouns in English and Korean. *Language and Information Society, 19*, 33-60.

Rao. D., & McMahon, B. (2022). *Natural language processing with PyTorch: Build intelligent language applications using deep learning*. O'Reilly.

Reuland, E. (2011). Syntax and interpretation systems: How is their labour divided? In C. Boeckx (Ed.), *The Oxford handbook of linguistic minimalism* (pp. 377-395). Oxford University Press.

Sperlich, D. (2020). *Reflexive pronouns: A theoretical and experimental synthesis (Vol.8)*. Springer.

Sang-Gu Kang, Assistant Professor
Department of English Language and Literature, Gangneung-Wonju National University
7, Jukheon-gil, Gangneung-si, Gangwon-do, Korea
E-mail: kangsg39@hanmail.net

Received: October 9, 2023
Revised: December 7, 2023
Accepted: December 8, 2023