

Sparks of Pure Competence in LLMs: the Case of Syntactic Center Embedding in English

Daniel Hardt
Copenhagen Business School
dha.msc@cbs.dk

Syntactic Embedding

Col Mustard killed Mr Boddy.

Col Mustard killed Mr Boddy in the library.

Col Mustard killed Mr Boddy in the library with the candlestick.

Col Mustard killed Mr Boddy in the library with the candlestick without remorse.

Left-periphery

Harry likes fish

John believes Harry likes fish

John believes Tom said Harry likes fish

John believes Tom said everyone knows Harry likes fish

Right-periphery

Level

0 The teacher is happy

1 The teacher the student saw is happy.

2 The teacher the student the driver hit saw is happy.

3 The teacher the student the driver the girl likes hit saw is happy.

4 The teacher the student the driver the girl the man hates likes hit saw is happy.

Center Embedding

Left and right peripheral embedding is unbounded

Center embedding at levels 2 or higher is generally uninterpretable and is exceedingly rare

Competence vs. Performance

According to the **Competence** Model, center embedding is fully grammatical. (Chomsky 1957)

Human **Performance** is different -- multiple levels of center embedding are unacceptable (Gibson 1998)

The Question

LLMs are not memorizing text data – they are learning some *about* the language data.

What are they learning? What *should* they be learning?

Is it human linguistic **Performance**, or is it **Competence**?

If an LLM is learning performance, it should fail with multiple center embeddings.

If it is learning competence, it should not.

Test 1

Q0: Question targets most deeply embedding clause

The teacher the student the driver saw hit is happy.

Q: Who saw who?

A: The driver saw the student.

GPT-4 remains highly accurate with 5 examples

Test 2

Q1: Question targets *next* most deeply embedding clause

The teacher the student the driver saw hit is happy.

Q: Who hit who?

A: The student hit the teacher.

GPT-4 has high accuracy up to level 3, with 5 examples

Test 3

Variable-length NPs:
NPs can be either one word or two words

The teacher the happy student the boy hit is happy.

GPT-4 needs 15 examples to reach high accuracy on level 3

Test 4

Missing VP Illusion
Gibson and Thomas (1999)

The teacher the student the driver saw hit is happy

vs.

The teacher the student the driver saw is happy

Missing VP

Conclusion

Unlike the other models, GPT-4 is more accurate than humans with multiple levels of center embedding

This suggests it is learning aspects of linguistic **Competence**, rather than **Performance**

The picture is a complicated one -- accuracy is sensitive to

- Level of embedding
- Number of examples
- Form of question

References

Noam Chomsky. 1957. Syntactic structures. The Hague: Mouton.

Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Edward Gibson and James Thomas. 1999. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3):225–248.