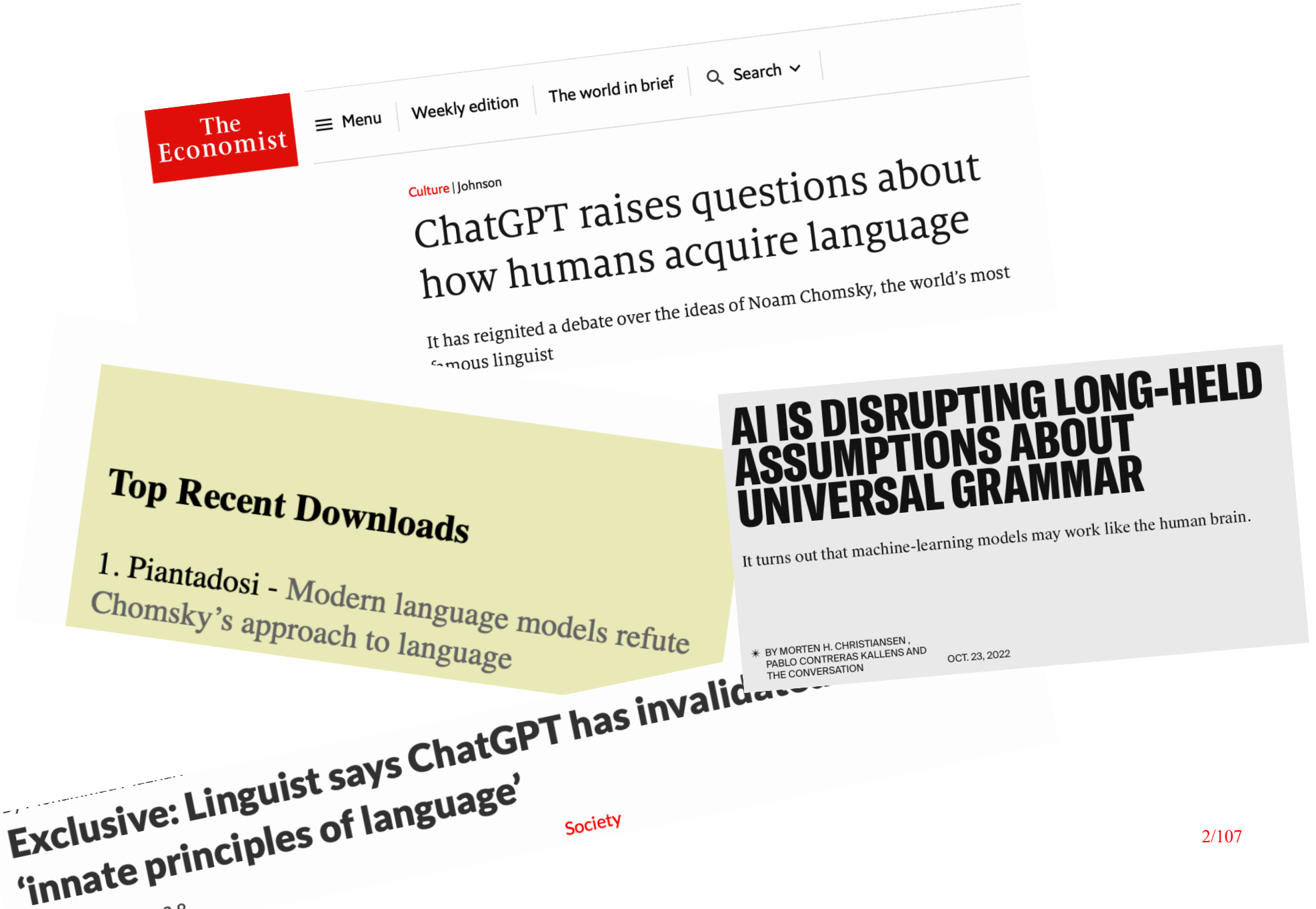
Is there an LLM challenge for Linguistics?
(or is there a Linguistics challenge for LLMs?)

David Pesetsky

MIT | Linguistics

The background of this presentation:



What I am: a linguist

What I am, more specifically: a *syntactician*

What I am, more specifically: a *syntactician* = a linguist who tries to uncover principles that underlie the syntax of the world's languages

What I am *not*: a specialist in computational linguistics or someone who has done sophisticated investigations of the capabilities of Large Language Models and chatbots based on them such as ChatGPT.

What I am *not*: a specialist in computational linguistics or someone who has done sophisticated investigations of the capabilities of Large Language Models and chatbots based on them such as ChatGPT.

.

Such people exist, but I'm not one of them.

My purpose today is

- to focus narrowly on whether an LLM-based device like ChatGPT, with the properties claimed for it (in particular, its data-drivenness), can *in principle* challenge what we think we know from linguistics about the nature of linguistic knowledge and how it is acquired —

  with an illustrative example that happens to be related to my own interests and my own work on syntax

- My projected audience: someone interested enough in linguistics to want to hear some linguistics in the context of today's talk, who is not a specialist in the field

- If *you* are a syntactician: my apologies for many corners to be cut and repetition of things you already know...

- My projected audience: someone interested enough in linguistics to want to hear some linguistics in the context of today's talk, who is not a specialist in the field

- If *you* are a syntactician: my apologies for many corners to be cut and repetition of things you already know...

- If you are a computer scientist and know something about LLMs: you surely know more than I do about the technology, so please correct me where I go astray.

I am not here as a Luddite:

- The engineering behind these systems may be of great use in different domains — others are better judges of that than me.

- And yes, the behavior of these systems is intrinsically fascinating and remarkable — and a near endless source of amusement and procrastination opportunities for amateurs like me.

I am not here as a Luddite:

- The engineering behind these systems may be of great use in different domains — others are better judges of that than us.
- And yes, the behavior of these systems is intrinsically fascinating and remarkable — and a near endless source of amusement and procrastination opportunities for amateurs like me.

- Even though the thing is very willing to declare me dead.

ChatGPT 3

PE

Who was David Pesetsky and when did he live?

David Pesetsky was a linguist and a professor of Linguistics at the Massachusetts Institute of Technology (MIT). He was born on December 31, 1955 and passed away on December 2, 2020. He was known for his research in syntax, the study of the structure of sentences in language.

ChatGPT 4

PE

You

Who was David Pesetsky and when did he live?

< 2 / 2 >

ChatGPT

David Pesetsky was a prominent American linguist known for his work in theoretical linguistics, particularly within the framework of generative grammar, a theory of grammar most famously associated with Noam Chomsky. He was born on August 15, 1957, and passed away on January 13, 2023. Pesetsky made significant contributions to the understanding of syntax, the part of linguistics that studies the structure of sentences. His work often focused on the innate structures of language and how these are reflected in various languages across the world. He was a professor at the Massachusetts Institute of Technology (MIT) and had a profound influence on the field of linguistics.

It is important to talk about these topics because of the expression of views like these in the public arena:

"*The rise and success of large language models undermines virtually every strong claim for the innateness of language that has been proposed by generative linguistics. Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes.*"

(Piantadosi 2023)

Modern language models refute Chomsky's approach to language

Steven Piantadosi

November 2023

The rise and success of large language models undermines virtually every strong claim for the innateness of language that has been proposed by generative linguistics. Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes. I describe the sense in which modern language models implement genuine theories of language, including representations of syntactic and semantic structure. I highlight the relationship between contemporary models and prior approaches in linguistics, namely those based on gradient computations and memorized constructions. I also respond to several critiques of large language models, including claims that they can't answer ``why'' questions, and skepticism that they are informative about real life acquisition. Most notably, large language models have attained remarkable success at discovering grammar without using any of the methods that some in linguistics insisted were necessary for a science of language to progress. (UPDATED: With a postscript on replies to the original draft)

Format: [pdf]

Reference: lingbuzz/007180
(please use that when you cite this article)

Published in:

keywords: large language model, minimalism, chomsky, generative syntax, emergent, computational modeling, statistical learning, cognitive science, syntax

previous versions:
v6 [October 2023]
v5 [September 2023]
v4 [March 2023]
v3 [March 2023]
v2 [March 2023]
v1 [March 2023]

Downloaded: 24461 times

Two months later:

Downloaded: 26611 times

It is important to talk about this topic because of the expression of views like these in the public arena:

"*The rise and success of large language models undermines virtually every strong claim for the innateness of language that has been proposed by generative linguistics.*

A short response: *no it doesn't*

Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes."

A short response: *no it hasn't*

Summary of the longer answer:

There is a logic and a structure to how the syntax of human languages works (and almost all other aspects of language as well) …

- which transcends the differences among various languages, i.e. the specific linguistic data to which individual children are exposed (the argument from *universals*)
- whose effects appear to be known to children despite the absence of relevant evidence to which children are actually exposed in their language-acquiring years (the argument from *poverty of the stimulus*)

… and which therefore cannot be explained purely as a result of learning from input, no matter how sophisticated and remarkable the system.

The evidence for this view is untouched by current claims about LLMs.

An illustrative example

An illustrative example

In English, a subordinate clause may be *large* …

- beginning with a *complementizer* such as *that*
- verb shows tense and agrees with the subject

a. We think [that Sue praised Mary] (shows tense)

b. We think [that Sue speaks French] (agrees with subject)

An illustrative example

In English, a subordinate clause may be *large* …

- beginning with a *complementizer* such as *that*
- verb shows tense and agrees with the subject

a. We think [that Sue praised Mary]

b. We think [that Sue speaks French]

… or *reduced in size* to varying degrees

- no *complementizer* such as *that*

c. We think [Sue praised Mary] (*that* missing)

- no tense, no agreement

d. I consider [Sue to speak French well] (tense and agreement missing)

An illustrative example

Crucial funny fact about English:

- If you question the subject of the subordinate clause, the subordinate clause *must* lack the complementizer — it is *obligatorily reduced in size*:

a. *Who do we think [that __ praised Mary]

b. ✓ Who do we think [__ praised Mary]

asterisk (*) = not grammatically well-formed

An illustrative example

Funny fact about English:

- If you question the subject of the subordinate clause the subordinate clause is obligatorily reduced in size — the complementizer must be missing …

a. *Who do we think [that __ praised Mary]

b. ✓ Who do we think [__ praised Mary]

- … but if you question the direct object (or anything else), there is no such effect:

c. ✓ Who do we think [that Sue praised __]

d. ✓ Who do we think [Sue praised __]

An illustrative example

Funny fact about English: called the ***that*-trace effect**

- If you question the subject of the subordinate clause the subordinate clause is obligatorily reduced in size — the complementizer must be missing …

a. *Who do we think [that __ praised Mary]

b. ✓ Who do we think [__ praised Mary]

- … but if you question the direct object (or anything else), there is no such effect:

c. ✓ Who do we think [that Sue praised __]

d. ✓ Who do we think [Sue praised __]

An illustrative example

Chat-GPT knows the *that*-trace effect and will correct errors …

Model: GPT-4

PE

Mary thinks that David Copperfield and Moby-Dick are the two most interesting English novels. Which novels does Mary think that are the two most interesting?

Mary thinks that the two most interesting English novels are "David Copperfield" and "Moby-Dick".

PE

Was my question grammatically well-formed?

Your question contained a slight grammatical error. The correct phrasing would be "Which novels does Mary think are the two most interesting?" The word 'that' after 'think' was unnecessary in this sentence. Here's how to phrase your original question more correctly:

"Mary thinks that 'David Copperfield' and 'Moby-Dick' are the two most interesting English novels. Which novels does Mary think are the two most interesting?"

An illustrative example

… presumably on the basis of patterns observed in texts

Model: GPT-4

PE

Mary thinks that David Copperfield and Moby-Dick are the two most interesting English novels. Which novels does Mary think that are the two most interesting?

Mary thinks that the two most interesting English novels are "David Copperfield" and "Moby-Dick".

PE

Was my question grammatically well-formed?

Your question contained a slight grammatical error. The correct phrasing would be "Which novels does Mary think are the two most interesting?" The word 'that' after 'think' was unnecessary in this sentence. Here's how to phrase your original question more correctly:

"Mary thinks that 'David Copperfield' and 'Moby-Dick' are the two most interesting English novels. Which novels does Mary think are the two most interesting?"

An illustrative example

- Does Chat-GPT's achievement tell us something about how human English speakers come to know the *that*-trace effect?

  or what it is and why it exists?

An illustrative example

- Does Chat-GPT's achievement tell us anything about how human English speakers come to know the *that*-trace effect?

  or what it is and why it exists?

- Here are two arguments suggesting that the answer is *no*:

Most of the linguistic discussion over the next ten minutes comes from:

Pesetsky, David. 2017. Complementizer-trace effects. In *Blackwell Companion to Syntax, 2nd Edition*. Oxford: Wiley-Blackwell. DOI: [:] 10.1002/9781118358733

Argument #1: from universality

The *that*-trace effect is found in languages all over the globe.

Argument #1: from universality

The *that*-trace effect is found in languages all over the globe

French (Perlmutter 1971, 99ff)

questioning the direct object:

a. Qui a-t-il dit [que Marie voulait voir __]?
who has-he said that Marie wanted to see
'Who did he say that Marie wanted to see?'

questioning the subject:

b. *Qui a-t-il dit [que __ voulait voir Marie]?
who has-he said that wanted to.see Marie
'Who did he say wanted to see Marie?'

Argument #1: from universality

Wolof [Senegal, Senegambian] (Martinović 2014)

questioning the direct object:

a. L-an l-a Aali xam [ni l-a xale bi gis __]
CM-Q l-CWH Ali know that *l*-CWH child DEF.SG see
'What did Ali know that the child saw'

questioning the subject:

b. K-an l-a Aali xam [ni mu (***l**-)a (>moo) __ gis xale bi\
CM-Q *l*-CWH Ali know that 3SG *l*-CWH see child DEF.SG
'Who did Ali know saw the child?'

Argument #1: from universality

Nupe [Nigeria, Atlantic-Congo] (Kandybowicz 2006, 220-221)

questioning the direct object:

a. Ke u: bè [ke Musa du __] na o?
what 3.SG seem COMP Musa cook NA O
'What does it seem that Musa cooked?

questioning the subject:

b. *Zèé u: bè [ke __ du nakàn] na o?
who 3.SG seem COMP cook meat NA O
'Who does it seem cooked meat?'

Argument #1: from universality

Levantine Arabic (Kenstowicz 1983; 1989)

questioning the direct object:

a. ʔayy fusṭaan [Fariid kaal (innu) l-bint ištarat _]
which dress Fariid said that the-girl bought
'Which dress did Fariid say that the girl bought?'

questioning the subject:

b. ʔayy bint Fariid kaal [(*innu) __ ištarat l-fusṭaan]
which girl Fariid said that bought the dress
'Which girl did Fariid say bought the dress?'

Argument #1: from universality

The generalization across languages:
Extracting the subject forces the clause to be smaller than full-sized.

Crucially: we do not find languages with the opposite pattern

— e.g. questioning a *non*-subject forcing the clause to be smaller

— or any other variant.

Argument #1: from universality

A generalization across languages:

Extracting the subject forces the clause to be smaller than full-sized.

Argument #1: from universality

- Even *complications* in the generalization hold across languages:

Adverb intervention effect (English) (Bresnan, 1977; Culicover 1993)

Which candidate did she think [that under these circumstances __ should be hired]?

Adverb intervention effect (Nupe) (Kandybowicz 2006)

Zèé Musa gàn [gànán pányi lèé __ nì enyà] o?
who Musa say that before PST beat drum O
'Who did Musa say that a long time ago beat the drum?'

Adverb intervention effect (French) (Bošković 2016)

?Quelle étudiante crois-tu [**que** dans deux jours __ va partir]?
which student believe-you that in two days goes leave.INF
'Which student do you believe that in two days is going to leave?'

Argument #1: from universality

- The fact that *languages in general* behave this way teaches us that someone's knowledge of the *that*-trace effect is unlikely to arise from that individual's exposure to data...

- ... precisely because it is a general fact about languages across the globe, not a fact particular to one speech community, much less one individual exposed to one particular corpus.

- Thesis:

  The effect is a by-product of some cognitive property that belongs to the species as a whole.

Argument #2: from research on acquisition

Evidence concerning the language acquisition process independently supports this conclusion.

Argument #2: from research on acquisition

11,308-utterance corpus of child-directed speech (Pearl & Sprouse, 2013)

- the crucial contrast is ✓(a) vs. *(c) …

	extraction type	**clause introducer**	**# of occurrences in 11,308 utterances**
a.	object	*that*	2
b.	object	---	159
c.	subject	*that*	0
d.	subject	---	13

Phillips (2013, 144)

- … but there is almost no difference in input frequency between these two conditions

Argument #2: from research on acquisition

- Corpora of *adult*-directed speech also analyzed by Pearl and Sprouse corpora contain a slightly greater percentage of object extractions from clauses introduced with *that*, but the overall number is still quite low

- "Even in the most 'helpful' corpus, the adult-directed speech corpus, we can estimate that the crucial object questions with overt *that* occur with sufficient frequency for a child to hear one roughly once every ten days"

(Phillips 2013)

Argument #2: from research on acquisition

- Of course, if LLM's input has a similar character, it might be dealing not with 2 vs. 0 relevant tokens — but 2 billion vs. some tiny number, a very different picture perhaps, and part of the explanation for why it "knows" the *that*-trace effect.

  But children are in a very different situation.

Argument #2: from research on acquisition

- Furthermore, there is direct experimental evidence of early knowledge of *that*-trace effect by children aged 2 years 10 months and up (Thornton 1990 and others)

Summarizing the arguments

- Universality: The *that*-trace effect is found in unrelated speech communities across the planet, and imaginable similar alternative effects are not.

- Poverty of the stimulus: While Chat-GPT 's corpus contains enough information to support learning the *that*-trace effect, a human child's corpus does not — and children reliably grow up to show knowledge of the effect.

- Thesis:

  The *that*-trace effect is a by-product of some property that belongs to the species as a whole.

Exceptions that prove the rule

Exceptions that prove the rule

- Though the *that*-trace effect is a universal insofar as it is found in unrelated languages all over the globe ...

- ... it is not a universal that every relevant language superficially appears to obey!

(by "*that*-trace effect" here I mean specifically a ban on extraction of the subject without deleting the complementiser)

Exceptions that prove the rule

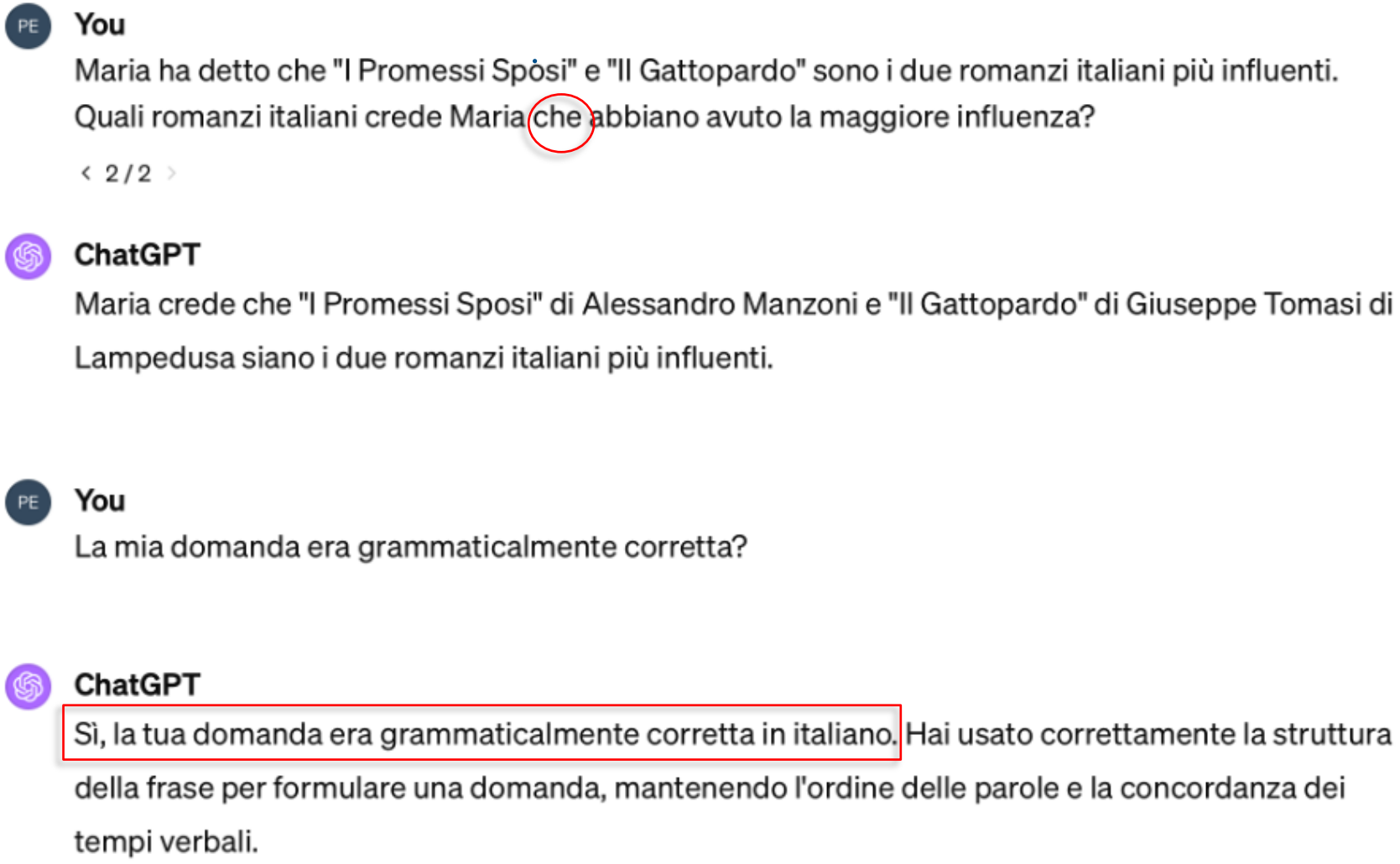
- Though the *that*-trace effect is a universal insofar as it is found in unrelated languages all over the globe ...

- ... it is not a universal that every relevant language appears to obey!

- and Chat-GPT is not ignorant of this!

Exceptions that prove the rule

Quali romanzi crede Maria **che** abbiano avuto la maggiore influenza?
what novels believes M. that have had the most influence
'What novels does Maria believe (that) have had the most influence?'



You

Maria ha detto che "I Promessi Spòsi" e "Il Gattopardo" sono i due romanzi italiani più influenti. Quali romanzi italiani crede Maria che abbiano avuto la maggiore influenza?

< 2 / 2 >

ChatGPT

Maria crede che "I Promessi Sposi" di Alessandro Manzoni e "Il Gattopardo" di Giuseppe Tomasi di Lampedusa siano i due romanzi italiani più influenti.

You

La mia domanda era grammaticalmente corretta?

ChatGPT

Sì, la tua domanda era grammaticalmente corretta in italiano. Hai usato correttamente la struttura della frase per formulare una domanda, mantenendo l'ordine delle parole e la concordanza dei tempi verbali.

Exceptions that prove the rule

ChatGPT 4

You

María piensa que "Don Quijote" y "Cien años de soledad" son las dos novelas en español más interesantes. ¿Qué novelas cree María que son las más interesantes?

ChatGPT

María cree que "Don Quijote" y "Cien años de soledad" son las dos novelas en español más interesantes.

You

¿Estaba mi pregunta gramaticalmente bien formulada?

ChatGPT

Sí, tu pregunta estaba gramaticalmente bien formulada. Fue clara y correctamente estructurada en español.

ChatGPT 4o

PE

Maria a spus că Enigma Otiliei și Bietul Ioanide sunt cele mai frumoase romane scrise în limba română. Care romane crede Maria că sunt cele mai frumoase dintre cele scrise în limba română?

Maria crede că „Enigma Otiliei” și „Bietul Ioanide” sunt cele mai frumoase romane scrise în limba română.

Întrebarea mea a fost corectă din punct de vedere gramatical?

Da, întrebarea ta a fost corectă din punct de vedere gramatical. Așa arată întrebarea corectă:

"Maria a spus că Enigma Otiliei și Bietul Ioanide sunt cele mai frumoase romane scrise în limba română. Care romane crede Maria că sunt cele mai frumoase dintre cele scrise în limba română?"

(thanks to Donca Steriade for the Romanian prompt)

1/2

Exceptions that prove the rule

(though on a second try, it preferred a different version entirely)

Da, întrebarea ta a fost corectă din punct de vedere gramatical. Totuși, pentru a suna puțin mai natural, ai putea reformula întrebarea astfel:

"Care sunt, în opinia Mariei, cele mai frumoase romane scrise în limba română?"

2/2

Exceptions that prove the rule

... and on the third try, it became quite annoying, I think ...

Întrebarea ta este corectă din punct de vedere gramatical, dar ar putea fi formulată puțin mai clar pentru a evita repetarea cuvântului „cele”. O variantă îmbunătățită ar putea fi:

„Maria a spus că Enigma Otiliei și Bietul Ioanide sunt cele mai frumoase romane scrise în limba română. Care romane crede Maria că sunt cele mai frumoase scrise în limba română?”

Astfel, întrebarea devine mai clară și mai fluidă.

3/3

Exceptions that prove the rule

Recall its English response:

Model: GPT-4

PE

Mary thinks that David Copperfield and Moby-Dick are the two most interesting English novels. Which novels does Mary think that are the two most interesting?

Mary thinks that the two most interesting English novels are "David Copperfield" and "Moby-Dick".

PE

Was my question grammatically well-formed?

Your question contained a slight grammatical error. The correct phrasing would be "Which novels does Mary think are the two most interesting?" The word 'that' after 'think' was unnecessary in this sentence. Here's how to phrase your original question more correctly:

"Mary thinks that 'David Copperfield' and 'Moby-Dick' are the two most interesting English novels. Which novels does Mary think are the two most interesting?"

Exceptions that prove the rule

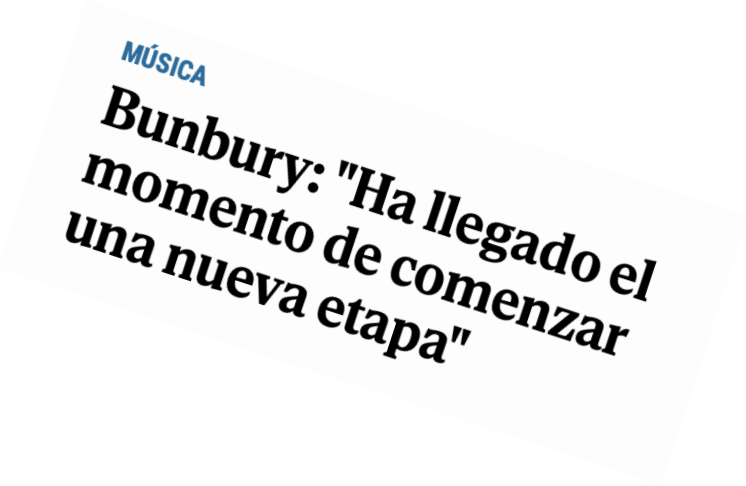
- What makes Italian, Spanish, and (I think) Romanian different from English and French ...

- ... such that these languages don't show the *that*-trace effect , but English and French do?

Exceptions that prove the rule

- What makes Italian, Spanish, and Romanian different from English and French ...

- ... such that they don't show the *that*-trace effect , but English and French do? [forgive the absence of Romanian examples]

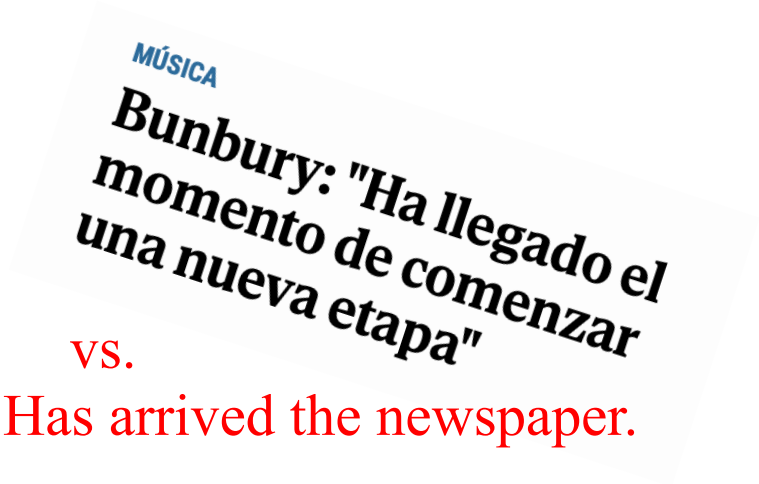


Doliu în Germania: a murit Karl-Heinz Schnellinger

Exceptions that prove the rule

- What makes Italian, Spanish, and Romanian different from English and French ...

- ... such that they don't show the *that*-trace effect , but English and French do?



vs.

*Has arrived the newspaper.

*Were invited to Bucharest several interesting people.

Doliu în Germania: a murit Karl-Heinz Schnellinger

Exceptions that prove the rule

Luigi Rizzi (1981 [updated]):

- The *that*-trace effect arises when a subject occupying a **verb-phrase-external position** is extracted from its clause.
- Languages that behave like Italian, but not English or French, permit the subject to alternatively occupy a **verb-phrase-*internal* position...**

Exceptions that prove the rule

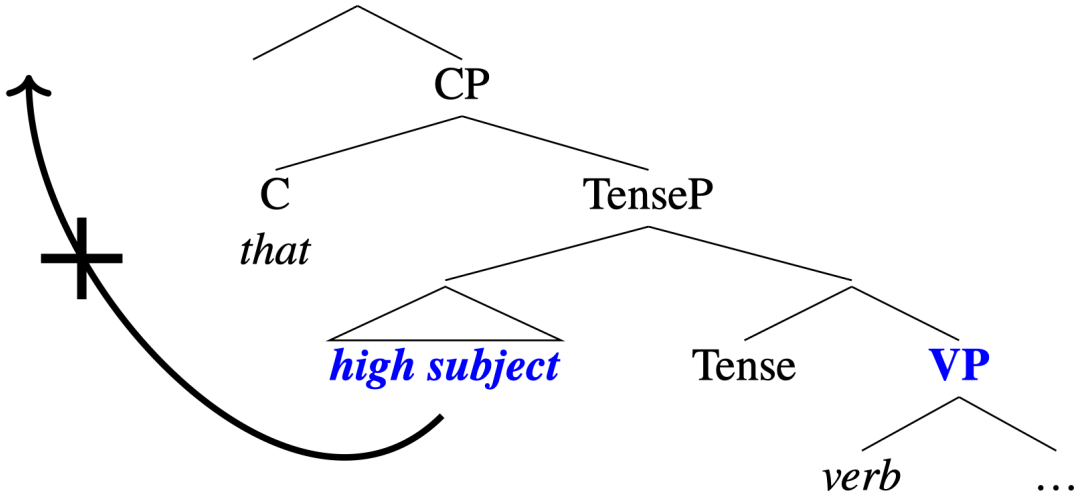
Luigi Rizzi (1981 [updated]):

- The *that*-trace effect arises when a subject occupying a **verb-phrase-external position** is extracted from its clause.
- Languages that behave like Italian, but not English or French, permit the subject to alternatively occupy a **verb-phrase-internal position...**
- ...And if extracted from that verb-phrase-internal position, there will be no *that*-trace effect.

  (Rizzi, in later joint work, calls this the "skipping strategy".)

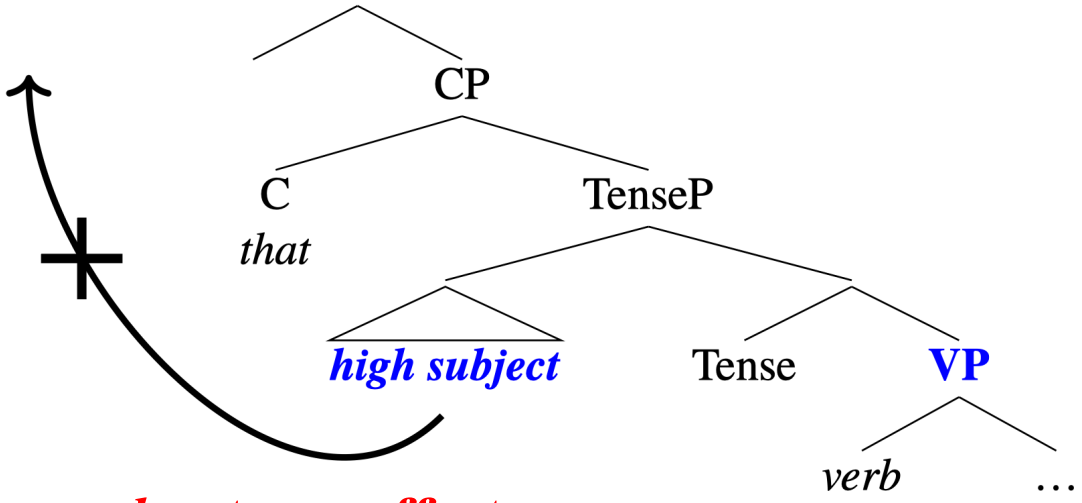
Exceptions that prove the rule

Producing a *that*-trace effect:

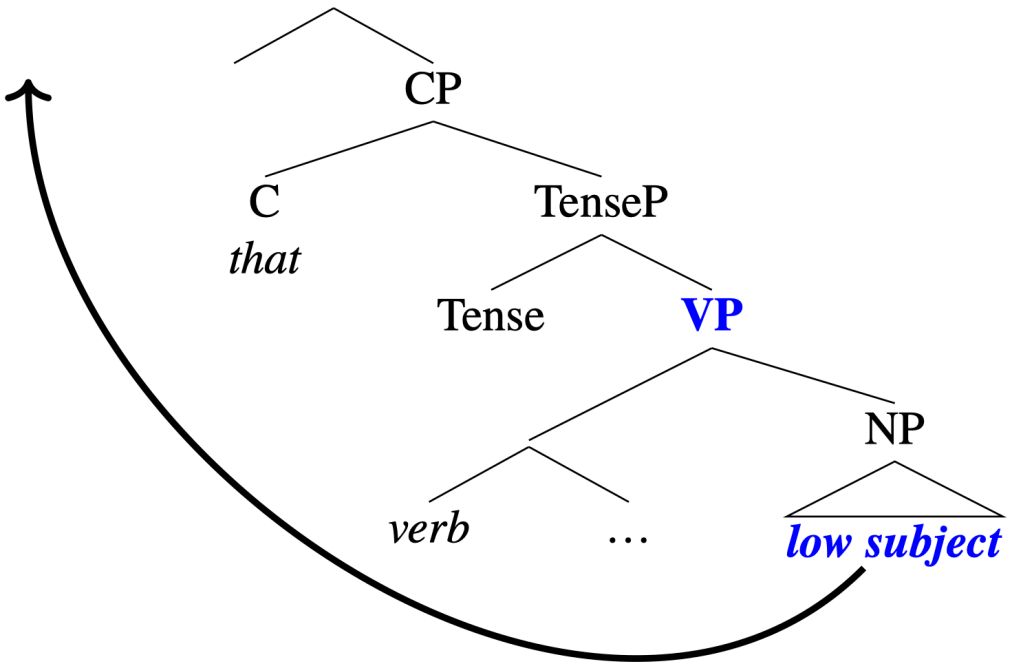


Exceptions that prove the rule

Producing a *that*-trace effect:



Avoiding a *that*-trace effect:



Exceptions that prove the rule

This is not a mere guess — it can be *shown* that languages like Italian truly do employ the "skipping strategy" to avoid the effect.

- Standard Italian - a pattern of partitive clitic pronoun use that diagnoses verb-phrase-internal vs. verb-phrase-external position (with *that*-trace violations showing the verb-phrase-internal pattern)

- Northern Italian dialects - a pattern of subject doubling that diagnoses verb-phrase-internal vs. verb-phrase-external position (with *that*-trace violations showing the verb-phrase-internal pattern)

Exceptions that prove the rule

(cf. (91)b). The obligatoriness of *ne* in (94) shows that (93)c is the only derivation available for (93)a, as our theory predicts:[38]

(94) Quante hai detto che *(ne) sono cadute?
"How many did you say that of-them-fell down?"

Consider also the following:

(95) (a) Quanti libri credi che siano usciti?
"How many books to you believe that have come out?"
(b) Quanti credi che *(ne) siano usciti?
"How many do you believe that of-them have come out?"

Given properties I–IV, the systematic obligatoriness of *ne* in these structures strongly supports our original hypothesis: the COMP-trace effect is circumvented in Italian by first moving the subject to postverbal position, and then *wh* extracting it from the clause.

The same holds for those questions where the Wh-subject is extracted from an embedded sentence, as in (41—42).

(41) a. Quante ragazze tu credi che e' sia venuto? (F)
b. Quante putele te pensi che sia vegnú? (T)
how-many girls you think that (it) has come
How many girls do you think have come?

(42) a. Quante ragazze tu credi che gli abbia parlato? (F)
b. Quante putele te pensi che abia parlá? (T)
how-many girls you think that (it) has spoken
How many girls do you think that (it) has spoken?

The grammatical sentences, where no agreement is shown, contrast with the corresponding sentences where the verb agrees with the Wh-phrase.

(43) a. *Quante ragazze tu credi che le siano venute? (F)
b. *Quante putele te pensi che le sia vegnude? (T)
how-many girls you think that they have come
How many girls do you think that they have come?

(44) a. *Quante ragazze tu credi che le abbiano parlato? (F)
b. Quante putele te pensi che le abia parlá? (T)
How many girls do you think that they have spoken?

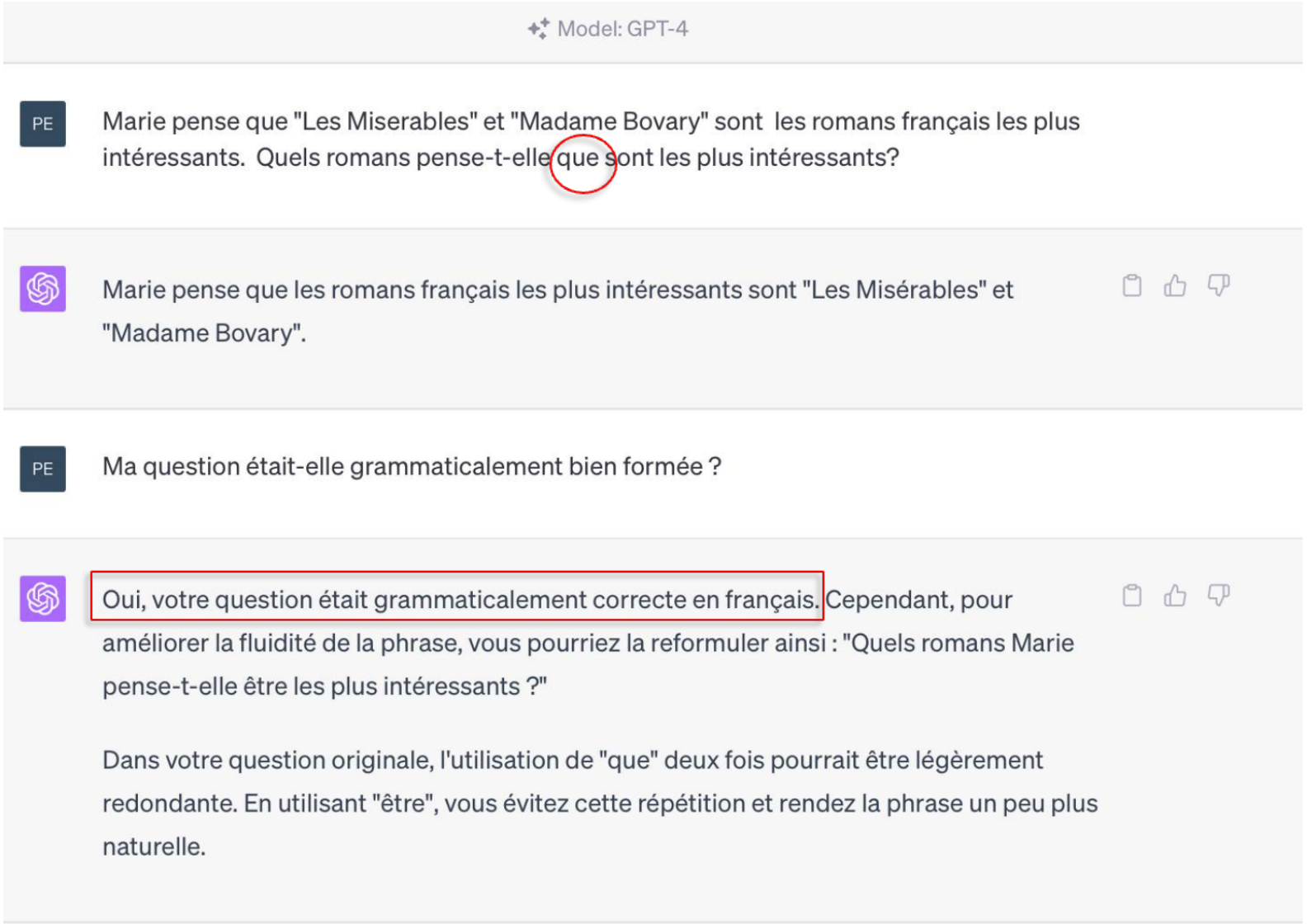
From the morphological characteristics of examples (36—38) and (41—42) we can unequivocally deduce that the Wh-phrase in each example originates in the postverbal position. Therefore the structure of these sentences can be represented by (45—46), respectively.

(45) a. [Quante ragazze$_i$ [*pro* gli $\left\{\begin{array}{l}\text{è venuto}\\ \text{ha parlato}\end{array}\right\}$ con te e_i]] (F)

b. [Quante putele$_i$ [*pro* $\left\{\begin{array}{l}\text{è vegnú}\\ \text{ha parlá}\end{array}\right\}$ con ti e_i]] (T)

Exceptions that prove the rule

By the way, ChatGPT actually gets a bit confused about French — or perhaps it's just being excessively polite?



Exceptions that prove the rule

No, it's actually confused:

PE

Mais je n'ai pas utilisé "que" deux fois.

Vous avez raison, je m'excuse pour cette erreur d'interprétation. Votre phrase était correcte et ne contenait qu'un seul "que".

Votre question initiale, "Marie pense que "Les Miserables" et "Madame Bovary" sont les romans français les plus intéressants. Quels romans pense-t-elle que sont les plus intéressants?" est correcte grammaticalement.

Je suis désolé pour la confusion. Merci de m'avoir corrigé.

Exceptions that prove the rule

The "skipping strategy" is not just about Southern Romance languages:

Bani-Hassan Arabic (Kenstowicz 1983; 1989)

high subject:

Fariid gaal innu al-binit istarat al-libaas.
F. said that the girl bought the dress.
'Fariid said that the girl bought the dress

low subject also possible:
Fariid gaal innu istarat al-binit al-libaas.
F. said that bought the girl the dress.

Exceptions that prove the rule

The "skipping strategy" is not just about Southern Romance languages:

Bani-Hassan Arabic (Kenstowicz 1983; 1989)

high subject:
Fariid gaal (innu) al-binit istarat al-libaas.
F. said that the girl bought the dress.
'Fariid said that the girl bought the dress

low subject also possible:
Fariid gaal (innu) istarat al-binit al-libaas.
F. said that bought the girl the dress.

no *that*-trace effect (due to skipping strategy):
wayy binit Fariid gaal (innu) istarat al-libaas?
which girl Fariid said that bought the dress
'Which girl did F. say bought the dress?'

Exceptions that prove the rule

Recall the contrast with Levantine Arabic (which does not permit low subjects):

Argument #1: from universality

Levantine Arabic (Kenstowicz 1983; 1989)

questioning the direct object:

a. ʔayy fusṭaan [Fariid kaal (innu) l-bint ištarat _]
which dress Fariid said that the-girl bought
'Which dress did Fariid say that the girl bought?'

questioning the subject:

b. ʔayy bint Fariid kaal [(*innu) __ ištarat l-fusṭaan]
which girl Fariid said that bought the dress
'Which girl did Fariid say bought the dress?'

Exceptions that prove the rule

Returning to the main line of argument...

- Let us imagine that Chat-GPT gets the *that*-trace effect or its absence right in every relevant language for which it has been exposed to a large enough corpus.

- But if it is truly just reproducing surface collocations found in the input data, it will miss the most important fact of interest to those of us who study human beings:

  the generalization discovered by Rizzi that ...

  ... it is languages that can place the subject in a low position within the clause, closer to direct objects, that permit it to be extracted with no *that*-trace effect.

Exceptions that prove the rule

- From a data-driven perspective, we might think the *that*-trace effect is just not a universal. Some languages do it, and some don't. Because the *that*-trace effect per se (absence of complementizer when there's a subject gap) is not present in Italian data.

- But it exerts its force in an entirely different way — forcing subjects to choose a low verb-phrase-internal position when they are going to be extracted. (And remember: *this has been shown*.)

  A linguistic *rule* is at stake, apparently universal or at least mysteriously ubiquitous — stated at a considerable level of abstraction.

Exceptions that prove the rule

But maybe Chat-GPT does know this after all? We don't really know what it knows, after all...

I'll return to that possibility shortly?

Are *that*-trace effects important — or just a "funny fact"?

We might discount the discussion so far as convincing but trivial.

Just a "funny fact" ...

Are *that*-trace effects important — or just a "funny fact"?

… a tiny speck of innate knowledge, perhaps, while the rest of language doesn't work that way?

Are *that*-trace effects important — or just a "funny fact"?

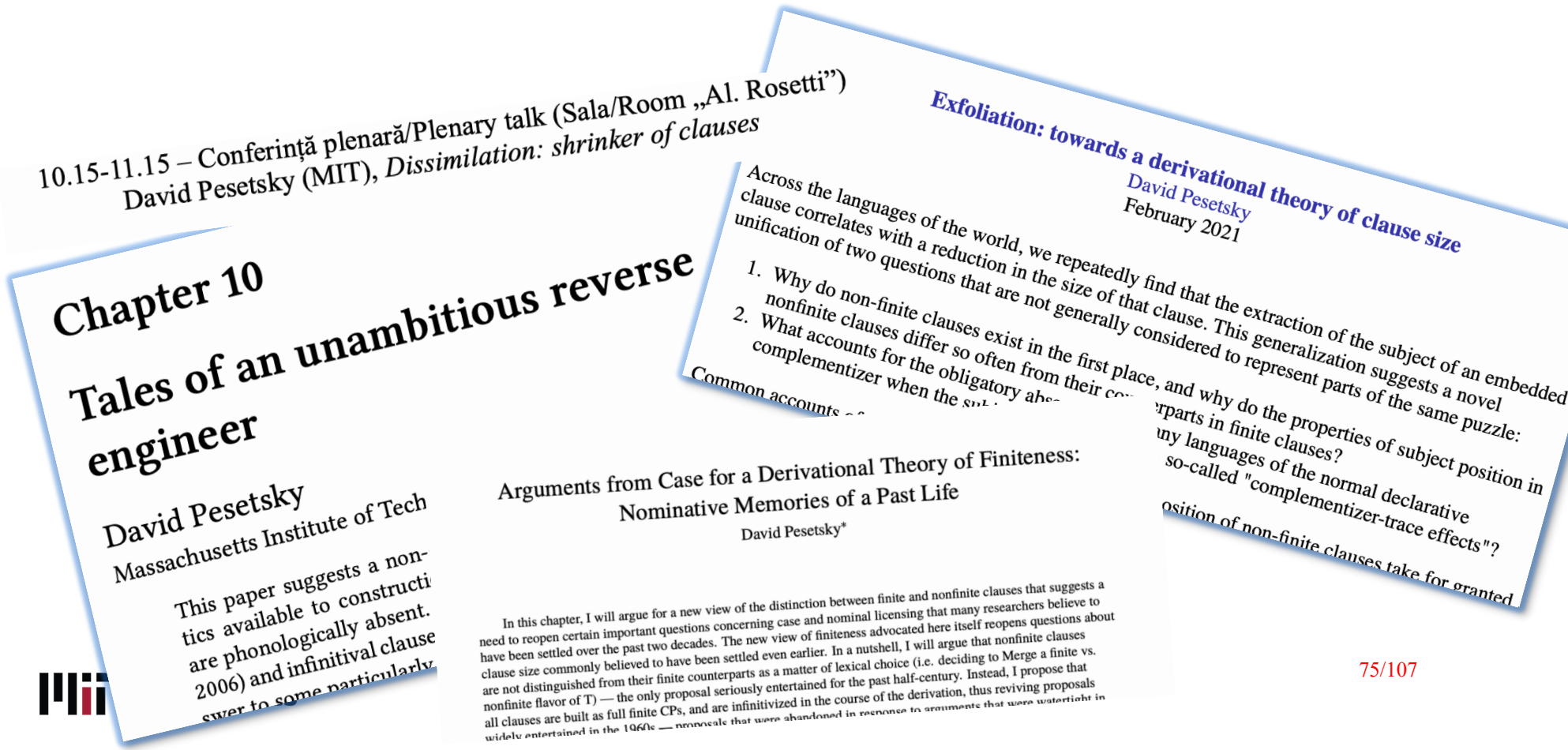
… a tiny speck of innate knowledge, perhaps, while the rest of language doesn't work that way?

As it happens, I've been working on this question (and will be discussing this at greater depth in my conference talk on Thursday)

Are *that*-trace effects important — or just a "funny fact"?

- My own research over the past decade or so suggests that *that*-trace effects are not just an isolated "funny fact"!

  This work argues that the broader generalization underlying the *that*-trace effect is a principal factor behind the multiplicity of clause types in the world's languages (including why clauses may also lack tense and agreement markers).



Are *that*-trace effects important — or just a "funny fact"?

The generalization underlying the *that*-trace effect is not specific to questions or other constructions that use *wh*-words.

- We are used to the idea that the subject of a clause bears some semantic relation to the predicate of the clause (agent of action, undergoer (= patient), experiencer ...

- But this is not always true: a rule of Raising to Subject available in many languages moves the subject of an embedded clause to form the subject of a higher clause

  — resulting in a subject that bears no semantic relation whatsoever to the main predicate of its clause.

Are *that*-trace effects important — or just a "funny fact"?

- Example from Lusaamia (Bantu, Kenya) [work of Carstens & Diercks 2013]:

Scenario: You find that the watering hole is empty. Though there are no cows on site, a speaker can truthfully say:

no raising to subject

(1) Bi-bonekhana [koti eng'ombe chi-ng'were amachi]
8SA-appear [that 10cow 10SA-drink 6water]
'It appears that the cows drank the water'

Are *that*-trace effects important — or just a "funny fact"?

- Example from Lusaamia (Bantu, Kenya) [Carstens & Diercks 2013]:

Scenario: You find that the watering hole is empty. Though there are no cows on site, a speaker can truthfully say:

no raising to subject

(1) Bi-bonekhana [koti eng'ombe chi-ng'were amachi]
8SA-appear [that 10cow 10SA-drink 6water]
'It appears that the cows drank the water'

raising to subject!

(2) Eng'ombe chi-bonekhana [__ chi-ng'were amachi]
10cow 10SA-appear [10SA-drink 6water]
'The cows appear to have drunk the water'

Are *that*-trace effects important — or just a "funny fact"?

- Example from Lusaamia (Bantu, Kenya) [Carstens & Diercks 2013]:

Scenario: You find that the watering hole is empty. Though there are no cows on site, a speaker can truthfully say:

no raising to subject

(1) Bi-bonekhana [koti eng'ombe chi-ng'were amachi]
8SA-appear [that 10cow 10SA-drink 6water]
'It appears that the cows drank the water'

raising to subject — with a *that*-trace effect!

(2) Eng'ombe chi-bonekhana [(*koti) __ chi-ng'were amachi]
10cow 10SA-appear [(*that) 10SA-drink 6water]
'The cows appear to have drunk the water'

(the facts are slightly more complicated than represented here)

Are *that*-trace effects important — or just a "funny fact"?

So the *that*-trace effect is not just about questions or similar constructions, but is a fact about multiple construction types.

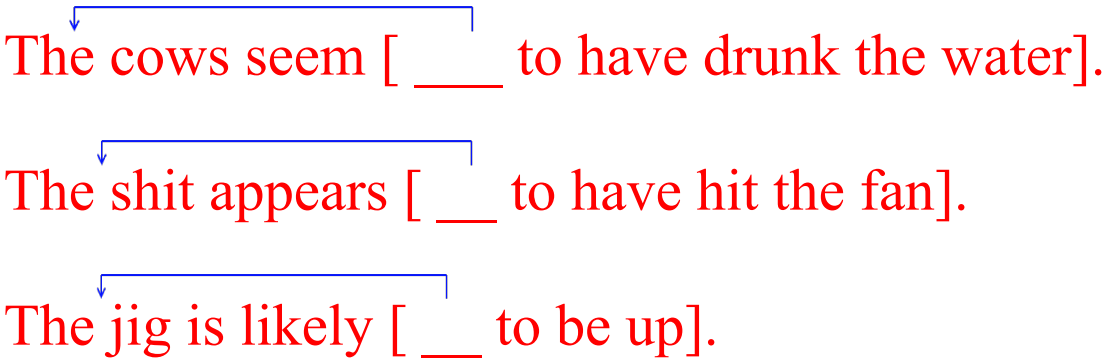
The law governing this effect is thus not specific to any single collocation of words. It is abstract, and structural.

But that's not all...

Are *that*-trace effects important — or just a "funny fact"?

- English has the construction as well — and just like Lusaamia, the complementizer must be absent ...

- ... and the tense and agreement morphology is also reduced (to *to*):



Are *that*-trace effects important — or just a "funny fact"?

- Of course Chat-GPT gets this right (though with annoying verbosity)

ChatGPT 4

PE **You**

You are in a tropical climate. You find that a watering hole is empty. Though there are no cows on site, can a speaker truthfully say "The cows appear that have drunk the water"?

ChatGPT

In a situation where a watering hole is empty and no cows are present, a speaker might use the phrase "The cows appear to have drunk the water" to suggest a hypothetical explanation for why the water is gone. This statement could be considered truthful under certain conditions:

1. **Previous Observations**: If there is evidence or prior knowledge that cows frequently visit this watering hole to drink and have been observed doing so before the water disappeared, the speaker could reasonably infer that the cows might have consumed the water.
2. **Tracks or Signs**: If there are physical signs that cows were recently present, such as hoof prints leading to and from the watering hole or cow droppings nearby, these could serve

Are *that*-trace effects important — or just a "funny fact"?

Of course Chat-GPT gets this right (though with annoying verbosity)

PE **You**

Was my question grammatically correct?

ChatGPT

Your original question, "The cows appear that have drunk the water," contains a slight grammatical error. Specifically, the phrase "appear that have" is not correctly structured in this context. The correct grammatical construction would be either:

- "The cows appear to have drunk the water," which implies that it seems the cows have drunk the water based on available evidence or inference.

or

- "It appears that the cows have drunk the water," which conveys a similar meaning but structures the sentence to explicitly state the appearance or inference before describing the action of the cows.

Both corrected versions provide a clearer and grammatically accurate way to express the inference that the cows are the reason the watering hole is empty.

Are *that*-trace effects important — or just a "funny fact"?

- More evidence that the *that*-trace effect works the same way for *wh*-question formation and Raising to Subject. Atypical subjects behave identically for the two effects in English (at least):

Locative inversion: the *that*-trace effect in questions (Bresnan 1972)

In this room can be found the best examples of Greek sculpture.

a. *In which room do you think [**that** __ can be found the best examples of Greek sculpture]?

b. ✓In which room do you think [__ can be found the best examples of Greek sculpture]?

Predicate inversion: the *that*-trace effect in questions

Even more important than syntax is climate change.

a. *How much more important than syntax do you think [**that** __ is climate change]?

b. ✓How much more important than syntax do you think [__ is climate change]?

Are *that*-trace effects important — or just a "funny fact"?

- More evidence that the *that*-trace effect works the same way for *wh*-question formation and Raising to Subject. Atypical subjects behave identically for the two effects in English (at least):

Locative inversion: the effect in Raising to Subject

In this room can be found the best examples of Greek sculpture.

[In this room] appear [__ to be found the best examples of ...]

Predicate inversion: the effect in Raising to Subject

Even more important than syntax is climate change.

[Even more important than syntax] seems [__ to be climate change]?

Are *that*-trace effects important — or just a "funny fact"?

- Once again, if I am right, there is a generalization behind all these facts that cut across the languages of the world and constructions within particular languages, which we can reformulate as follows:

Extracting the subject diminishes the complementizer or the verbal morphology of its clause (creating an infinitive), ***or both***.

(Of course, we need to make precise the conditions under which the complementizer vs. verbal morphology is diminished.)

Are *that*-trace effects important — or just a "funny fact"?

- And if this generalization is on the right track, it suggests that *that*-trace effects are not an isolated "funny fact", but a special case of a broader cross-linguistic generalization that predicts the existence of reduced clauses — a phenomenon that pervades the everyday utterances of most or at least many of the world's languages.

Are *that*-trace effects important — or just a "funny fact"?

- Of course, this is where a talk for syntax specialists would *begin*. Today, it is where I end.

For any specialists not coming to my talk tomorrow:

1. The generalization is actually about local movement of the subject to the clause periphery, the necessary first step of any long-distance movement process. So it is not even just about extracting a subject from its clause, but actually about highly local movement that precedes long-distance movement.

2. Cases that fall under the generalization:
   - anti-agreement (found under highly local wh-movement)
   - complementizer deletion and alteration under local wh-movement of subjects (examples from French, Wolof, Bùlì)
   - obligatory control as analyzed by Chierchia and extended by Landau: very local movement of a semantically vacuous pronoun, creating a predicate

Are *that*-trace effects important — or just a "funny fact"?

- But even if that specific work of mine is wrong …

  … very similar discussions to this one could be had about almost every aspect of grammatical structure.

Conclusions

Conclusions

There is a logic and a structure to how the syntax of human languages works …

- which transcends the differences among various languages, i.e. the specific linguistic data to which individual children are exposed (the argument from *universals*)
- whose effects appear to be known to children despite the absence of relevant evidence to which children are actually exposed in their language-acquiring years (the argument from *poverty of the stimulus*)
- which is deep and abstract, transcending the details of particular constructions and word collocations (the argument from *deep generalizations*)

Conclusions

If LLMs also appear to "know" about these generalizations that cut across languages and constructions, acquired with practically no evidence …

- **and if they were indeed good models of humans** — it would have to mean that something about their inherent architecture really does incorporate innate constraints on how grammars work, contrary to claims about how they were built and what they are doing. *Unlikely, but unbelievably exciting if true: the linguistics challenge for machine systems!*

- **More likely: they are not modeling human linguistic abilities at all** — but responding in a very different way from human children to the extraordinarily *non*-impoverished data to which they have been exposed, producing a simulacrum of linguistic competence but not the real thing. My guess: *this is probably the right answer.*

Conclusions

- **The point is: there is an intrinsic nature to human language *as a whole*.**

  Like LLMs, children do learn from data to which they are exposed — that's what it means to say "I speak English" and "I do not speak Nupe". But their learning takes place within a space of possibilities pre-defined by this intrinsic nature.

- It could be exciting to collaboratively develop and explore machine-implementable tools that model what we *actually* know about human language — and it's rather tragically weird that such collaborations hardly exist.

Conclusions

- A device that stitches together fragments of text from a huge corpus on the basis of probabilities is *not* a model of human behavior.

Conclusions

For example, the most recent version of ChatGPT is reluctant to kill me off, but shows a striking willingness to "hallucinate" a birthday for me — presumably picking probable dates when it hasn't learned the real one:

Conclusions

David Pesetsky is an American linguist known for his work in syntax and morphology. As of my last update in 2023, David Pesetsky is alive and continues to contribute to the field of linguistics. He was born on September 8, 1957. Pesetsky has held significant academic

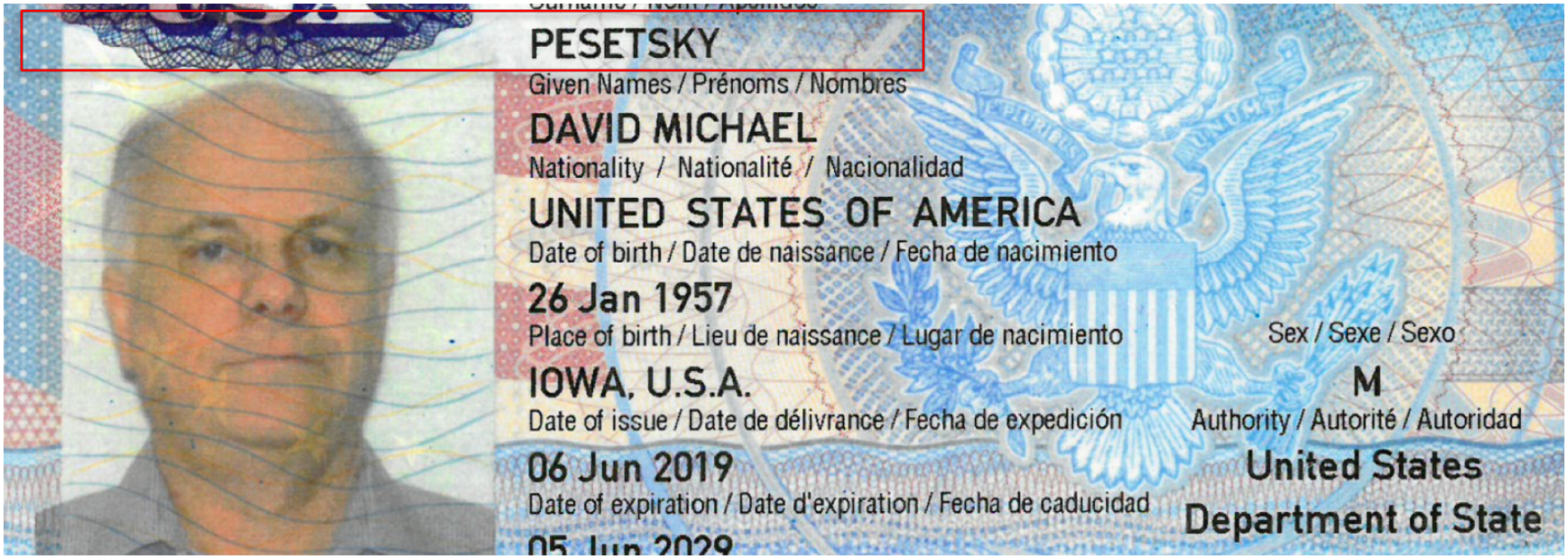
Conclusions

David Pesetsky is a prominent American linguist known for his contributions to syntactic theory and the study of language acquisition. Born on November 30, 1957, he has been a professor at the Massachusetts Institute of Technology (MIT) and has significantly

Conclusions

David Pesetsky is an influential American linguist born on November 14, 1957. He is well-known for his work in syntax and morphology, particularly in the areas of phrase structure and the theory of argument structure. As of my last update in 2023, David Pesetsky is a

Conclusions



Conclusions

- There is lots of concern in the press and in internet chatter about whether the descendents of ChatGPT and other LLMs will bring about the extinction of the human species, permit authoritarians to win elections, end all life on earth, and so forth...

Conclusions

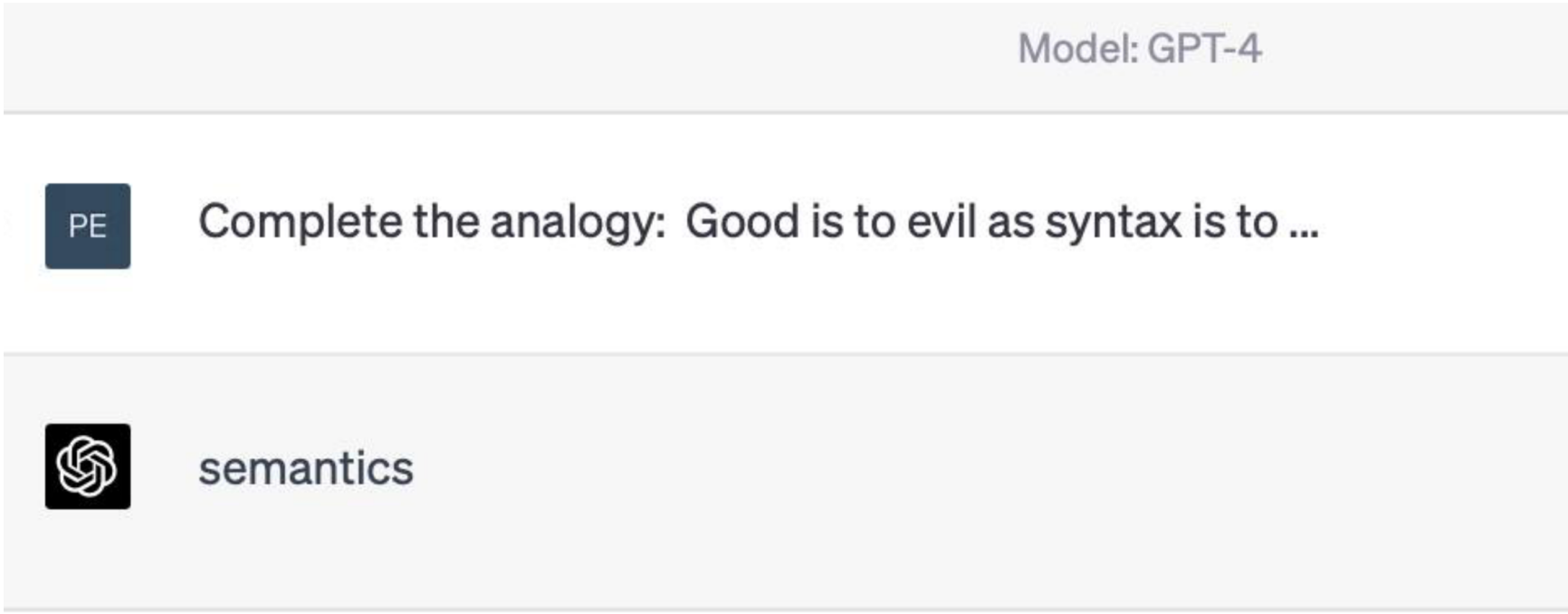
- There is lots of concern in the press and in internet chatter about whether the descendents of ChatGPT and other LLMs will bring about the extinction of the human species, permit authoritarians to win elections, end all life on earth, and so forth...

- ... things we are already doing quite well, without LLMs helping us, in fact.

Conclusions

- Though there is a delicious malevolence to some of ChatGPT's output

Conclusions



Conclusions

- My goal in this talk has been much more modest — to show how human language differs in principle in its deepest essence from what ChatGPT and similar systems do with text.

  Not an original goal either — but one that needs a bit more explanation to a general educated public than was provided by Chomsky and colleagues in their widely publicized *New York Times* commentary:

The New York Times

OPINION

GUEST ESSAY

Noam Chomsky: The False Promise of ChatGPT

March 8, 2023

But ChatGPT and similar prog[illegible] are, by design, unlimited in what they can "learn" (which is to say, memorize); they are incapable of distinguishing the possible from the impossible. Unlike humans, for example, who are endowed with a universal grammar that limits the languages we can learn to those with a certain kind of almost mathematical elegance, these programs learn humanly possible and humanly impossible languages with equal facility. Whereas humans are limited in the kinds of explanations we can rationally conjecture, machine learning systems can learn both that the earth is flat and that the earth is round. They trade merely in probabilities that change over time.

By Noam Chomsky, Ian Roberts and Jeffrey Watumull

Conclusions

Finally, here's a short version of this talk: summarized by ChatGPT itself

ChatGPT 4

You

Steven Piantadosi has written this:

"The rise and success of large language models undermines virtually every strong claim for the innateness of language that has been proposed by generative linguistics. Modern machine learning has subverted and bypassed the entire theoretical framework of Chomsky's approach, including its core claims to particular insights, principles, structures, and processes."

Is he right? Just answer yes or no.

ChatGPT

No.

Conclusions

Thank you!