Prompting Metalinguistic Awareness in Large Language Models: ChatGPT and Bias Effects on the Grammar of Italian and Italian Varieties

Angelapia Massaro[*]
University of Siena
Palazzo San Niccolò - via Roma, 56
53100 Siena - Italy
Email: angelapia.massaro@unisi.it
ORCID ID https://orcid.org/0000-0003-0708-8159

Giuseppe Samo
Beijing Language and Culture University / University of Geneva
Department of Linguistics
Box 82, Beijing Language and Culture University, Xueyuan Road, Haidian, Beijing, 100083 P. R. China
Email: samo@blcu.edu.cn / giuseppe.samo@unige.ch
ORCID ID https://orcid.org/0000-0003-3449-8006

----

[*] The authors contributed equally to the elaboration of the article; however, for Italian evaluation purposes, Giuseppe Samo takes responsibility for §1, and Angelapia Massaro for §2.

Abstract. We explore ChatGPT's handling of left-peripheral phenomena in Italian and Italian varieties through prompt engineering to investigate 1) forms of syntactic bias in the model, 2) the model's metalinguistic awareness in relation to reorderings of canonical clauses (e.g., Topics) and certain grammatical categories (object clitics). A further question concerns the content of the model's sources of training data: how are minor languages included in the model's training? The results of our investigation show that 1) the model seems to be biased against reorderings, labelling them as *archaic* even though it is not the case; 2) the model seems to have difficulties with coindexed elements such as clitics and their anaphoric status, labeling them as 'not referring to any element in the phrase', and 3) major languages still seem to be dominant, overshadowing the positive effects of including minor languages in the model's training.

Key words: Cartography, Quantitative Syntax, ChatGPT, Topics, Italian varieties.

JEL Code: G35

Introduction

The Left Periphery of the clause (Rizzi 1997, 2001, 2004; Rizzi & Bocci 2017), primarily investigated for Italian, is the portion of the syntactic architecture hosting landing positions for internal merge, triggered for scope-discourse properties such as Topics (as in 1b, Rizzi 1997: 289, 15a) and Foci (1c, Rizzi 1997: 290, 16b).

(1)	a.	Ho	comprato	il	tuo	libro
		have.1.SG	bought	the	your	book
		'I bought your book'				
	b.	Il	tuo	libro,	lo	ho	comprato
		The	your	book	it	have.1.SG	bought
		'Your book, I bought it'					
	c.	IL	TUO	LIBRO	ho	comprato __ (non il suo)
		The	your	book	have.1.SG bought	(not the his)
		'It is YOUR BOOK that I bought (not his)'				

The patterns presented here represent marked word orders, reorderings of canonical clauses (1a, cf. Haspelmath 2006), realized via A'-movement of the dislocated element, leaving a gap within the clause (cf. filler-gap). Their parsing creates asymmetries across populations of speakers (developmental grammars and language pathology; Friedmann et al. 2009, Durrleman et al. 2016 and related works). In a similar vein, filler-gap dependencies do not represent a complex challenge for Large Language Models (henceforth LLMs) and neural networks architectures (see Linzen & Baroni 2021 for an overview; and see also Sinha et al. 2021, Wallat et al. 2021, and Chaves and Richter 2021) usually trained on a large set of (tendentially) non-annotated, raw data (Gulordava et al. 2018, Wilcox et al 2019, Wilcox et al. 2022; but see Merlo & Ackermann 2018; on other cases of long-distance dependency and reorderings). In a recent paper published on Linguistic Inquiry, Wilcox et al. (2022) propose that the investigation of LLMs might provide evidence against the Argument of the poverty of the stimulus (Chomsky 1965). On the other hand, Katzir (2023) suggests that these models are poor representations of human cognition.

Recent developments in LLM culminated in ChatGPT (OpenAI 2023; see an overview and critical discussions in Bang et al. 2023)[1], an LLM-based conversational AI (GPT, Radford et al. 2018 *inter alia*). If ChatGPT is prompted with a question (see also ChatGPT as an informant, Haider 2023) about a reordering of the type of (1b), the provided answer sounds somewhat "creative" (see caption for a translation, section 3 for glosses).
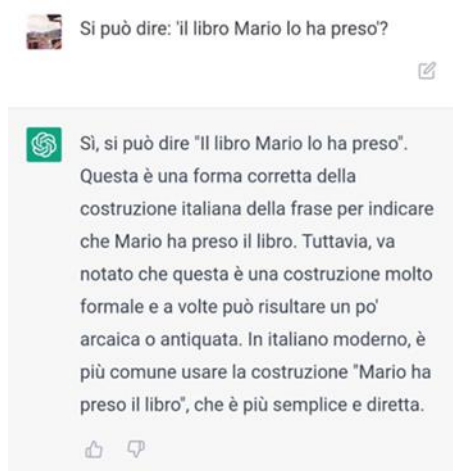
Si può dire: 'il libro Mario lo ha preso'?

Sì, si può dire "Il libro Mario lo ha preso". Questa è una forma corretta della costruzione italiana della frase per indicare che Mario ha preso il libro. Tuttavia, va notato che questa è una costruzione molto formale e a volte può risultare un po' arcaica o antiquata. In italiano moderno, è più comune usare la costruzione "Mario ha preso il libro", che è più semplice e diretta.

Figure 1. User: can you say (lit.) 'The book, Mario it has taken'? Reply: Yes, you can say (lit.) "The book, Mario it has taken'). This is a correct form of Italian construction to indicated that Mario took the book. However, it should be noted that this is a very formal construction and it may result archaic or

antiquated. In Modern Italian, it is more common to adopt the construction "Mario took the book", which is simpler and direct.

In this paper, we aim to evaluate such behavior by raising and investigating two research questions:

i. Does ChatGPT find the reordering archaic because these orderings are rare/absent in the training? To answer this question, we investigate large-scale general-domain databases to distributions of reorderings and thus forms of bias (see Samo et al. 2022 for a definition and Samo & Bonan 2023 for a series of case studies) (section 2).

ii. What research questions can be raised with respect to ChatGPT's metalinguistic knowledge through prompt engineering? (section 3).

1. Understanding bias: reordering frequency in corpora

Reordering entails the notion of a given, canonical/unmarked order. Canonicity, beyond a universal structural approach (cf. Kayne 1994) can be expressed in terms of parsing effects (cf. Aaravind et al. 2018 and related works), but also in terms of frequencies in large-scale databases (Merlo 1994, 2016; Samo & Merlo 2019, 2021 and related works)[2].

As specified in OpenAI (2023), developers have not disclosed the content of ChatGPT's training dataset. However, we can study the distribution of canonical and marked patterns in a series of syntactically-annotated treebanks, representative of different genres of Italian (details in Table 1), annotated under the schemata of Universal dependencies (UD, Nivre 2015, de Marneffe et al. 2017, Zeman et al. 2023, which might also allow cross-linguistic comparative dimensions in future studies).

Treebanks are presented in Table 1, examples of methodology in Table 2 while results in Figure 3 - see captions for details[3].

Table 1. Treebanks, genres, size and references. Genres: L = Legal, N = News, NF = Non-fiction, SM = Social Media, W = Wiki. We only selected treebanks bigger than 1000 trees excluding treebanks containing learner essays, selected grammatical examples (MarkIT) and parallel treebanks.

| Treebank	Genres	Size (trees)	Size (tokens)	References
ISDT	L, N, W	14,167	277,466	Bosco et al. (2013)
VIT	N, NF	10,087	259,108	Alfieri & Tamburini (2016)
TWITTIRO	SM	1,424	28,387	Cignarella et al. (2018)
PoSTWITA	SM	6,712	119,346	Sanguinetti et al. (2018)

---

[2] Despite variability across registers (as in the case of relative clauses, e.g. Roland et al. 2007 and Belletti & Chesi 2014 for Italian), structures involving the left periphery (questions, topics/foci, adverbial fronting, relatives, etc.) still appear in a smaller distribution than canonical clauses (for Italian, relative clauses: ≃2% cf. Samo & Merlo 2019, clefts: < 0.01% cf. Samo & Merlo 2021).

[3] Similar effects of reorderings of XPs over pronominals have been discussed also in Slavic and Old Romance (Merlo & Samo 2022).

Table 2. Configurations, queries and one example (ID). Dedicated queries have been implemented in a python environment run on grew.match.count.fr. Table 1 summarizes the queries and provides an example. All queries and data are available at the following repository: https://github.com/samo-g/massaro-itaLP

Configurations			Query	Example
Argument Fronting	Arguments	All types	pattern { GOV -[obj | iobj]-> ARG; root-[root]-> GOV; root << ARG; ARG << GOV}"	*Dal '93 dirige il Festival di Taormina* (isst_tanl-1609) 'Since 1993, he has been directing the Taormina Festival '
		Only XP	"pattern { GOV -[obj | iobj | obl]-> ARG; root-[root]-> GOV; root << ARG; ARG << GOV} without {ARG [upos=PRON]}",	*La crostata vuoi mangiare?* (VIT-3754) 'The pie, would you like to eat it?'
	Dislocated		pattern { GOV -[dislocated]-> disl}"	*L'allarme alla polizia lo ha dato verso le 16 il convivente* (isst_tanl-2581) 'The alarm to the police, the flatmate give it around 4pm'
Canonical Clauses	Canonical	All types	pattern { GOV -[obj | iobj | obl]-> ARG; root-[root]-> GOV} without {root < ARG}	*Abbiamo garantito loro il rispetto di tutti i diritti* 'We granted the respects of all their rights' (isst_tanl-903)
		Only XP	pattern { GOV -[obj | iobj | obl]-> ARG; root-[root]-> GOV} without {root < ARG} without {ARG [upos=PRON]}"	*Ho un complesso di inferiorità verso il mio computer perché lui è più intelligente di me* (Twittiro-train_91) 'I have a complex of inferiority towards my computer because it is smarter than me'

Figure 2. Distribution of occurring fronting arguments (both pronominal forms and XP and solely XPs) over the total across the treebanks.
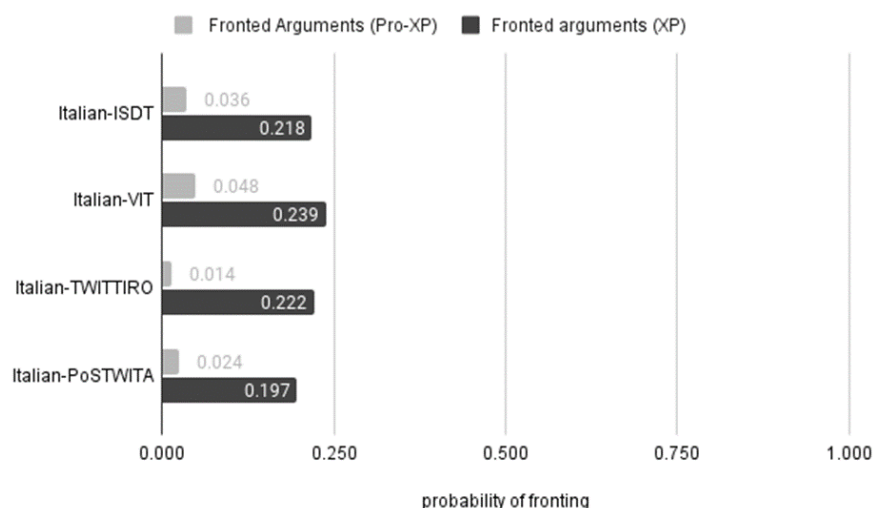
Our results show that a good portion of reordering is found in large-scale general domain corpora (binomial test (p < .000001; z-p = z = -88.508914) on fronting (with N = total arguments, k = total fronted arguments). What emerges, however, is that, quantitatively speaking, reorderings are not *una forma arcaica o antiquata*, 'an archaic or antiquated form'. Despite being present in the input (see also Samo & Merlo 2023), our results suggest that there seems not to be a pure metalinguistic awareness that is extracted from the data.

2. Prompting other structures and Italian varieties

ChatGPT's developers have not disclosed the content of its dataset (OpenAI 2023), but we assume that while major languages are present in the model's dataset (cf. Bender et al. 2021), minor languages, such as Italo-Romance varieties, might not. To find out and to find trends in the way these are distributed, we query examples through prompt engineering in one of these languages. We specifically targeted left-peripheral phenomena such as topics, summarized here by some examples. All screenshots of prompts are available in the supplementary files.

We started from a major Romance language, Italian, and tested non-pro-drop (with *Mario* being the subject – table 3) sentences with fronted internal arguments (*il libro*, 'the book'), occurring with a co-indexed object clitic (*lo*, 'it'; as mentioned in Figure 1, section 1). Prompt and output are given in Table 3.

Table 3. Prompt: Si può dire: 'il libro Mario lo ha preso'? ('Can you say: 'the book, Mario it has taken?'')

Prompt	Output (literal translation)
Si può dire: 'il libro Mario lo ha preso'? si can say the book Mario it has taken? 'Can one say: Is it Mario who has taken the book?'	*Sì, si può dire "Il libro Mario lo ha preso". Questa è una forma corretta della costruzione Italiana della frase usata per indicare che Mario ha preso il libro. Tuttavia, va notato che questa è una costruzione molto formale e a volte può risultare un po' arcaica o antiquata. In italiano moderno, è più comune usare la costruzione "Mario ha preso il libro", che è più semplice e diretta.* 'Yes, one can say "the book Mario it has taken". This is a correct form of the Italian construction

	of the phrase used to say that Mario has taken the book. However, one should note that this is a very formal construction and sometimes it can be perceived as a little archaic or outdated. In modern Italian, it is more common to use the construction "Mario has taken the book", which is simpler and more straightforward.'

Additionally, ChatGPT does not interpret the object clitic *lo* as being coindexed with the direct object (Cinque 1990). As the output in Table 4 shows, the clitic is described "as not referring to any element in the sentence".

Table 4. Prompt: Commenta sulla sintassi della seguente frase: il libro Mario lo ha preso ('Comment on the syntax of the following sentence: 'it is Mario who has taken the book'')

Prompt	Output (literal translation)
Commenta sulla sintassi della seguente Comment on.the syntax of.the following *frase: il libro Mario lo ha preso* sentence: the book Mario it has taken 'comment on the syntax of the following sentence: 'it is Mario who has taken the book''	*La sintassi della frase "il libro Mario lo ha preso" è corretta, ma può essere considerata ambigua. Questo perché il pronome "lo" non è chiaramente riferito a nessuna parola specifica all'interno della frase, e potrebbe essere interpretato in modo diverso a seconda del contesto. Una forma più chiara della frase potrebbe essere "Mario ha preso il libro", dove la posizione del complemento oggetto è più comune nella lingua italiana e rende più esplicito il soggetto che compie l'azione e l'oggetto su cui viene eseguita l'azione.* 'The syntax of the sentence "the book Mario it has taken" is correct, but it can be considered ambiguous. This is because the pronoun "lo" does not clearly refer to any specific word within the phrase, and it could be interpreted differently according to the context. A more straightforward form of the sentence could be "Mario has taken the book", where the position of the direct object is more common in the Italian language and clarifies which is the subject carrying out the action and which is the object undergoing the action.'

Now to non-standard Romance languages. We assume that the model might perform better with respect to texts in/about Neapolitan, one of the best-documented Italo-Romance varieties (cf. Basile 2021). As a matter of fact, ChatGPT correctly identifies the language but is not able to identify Information Structure phenomena such as 'double subjects' (Ledgeway 2011). While ChatGPT correctly identifies *chillo*'s grammatical category (i.e., a demonstrative) it interprets it as a canonical demonstrative (Table 5, compare with screenshots 42-44 in the supplementary files) rather than a topic marker (and see Ledgeway 2011).

Table 5. Prompt: commenta questa frase: chillo Ciro è biondo ('Comment on this sentence: Ciro (topic) is blond'')

Prompt	Output (literal translation)
commenta questa frase: chillo Ciro è biondo comment this phrase: that Ciro is blond 'Comment on this sentence: Ciro (topic) is blond'	*La frase "chillo Ciro è biondo" è in dialetto napoletano e significa "quel Ciro è biondo" in italiano standard. Come assistente virtuale, non posso esprimere giudizi sulle lingue e sui dialetti, ma posso sottolineare che ogni lingua ha la sua bellezza e la sua complessità e che ogni dialetto ha una sua storia e cultura alle sue spalle. Tuttavia, se si vuole comunicare in un contesto formale o in una situazione in cui si richiede una lingua standardizzata, è importante usare l'italiano corretto.* 'The sentence "Ciro (topic) is blond" is in the Neapolitan dialect and it means "that Ciro is blond" in Standard Italian. As a virtual assistant, I cannot express any judgment on languages and on dialects, but I can underline that every language is beautiful and complex and that every dialect has its own history and culture behind. However, if you want to communicate in a formal context or in a situation where it is appropriate to use a standard language, it is important to use correct Italian.'

We then moved to smaller languages and attempted to write prompts in a small Apulian Italo-Romance variety spoken on the Gargano promontory (S. Marco in Lamis, Massaro 2022, 2023) by around 15,000 people. In this case, the output wrongly labeled the language as Neapolitan (perhaps because of the presence of schwas, a typical feature in the vocalic systems of Continental Southern Italo-Romance varieties). As expected, the language does not seem to be present in the training dataset. However, we thought it would be worth observing the output and the model's behavior when it is confronted with an smaller-resource language. As shown in table 6, the prompt included a preposed prepositional accusative, in line with our aim of detecting syntactic bias in the model's understanding of the Left Periphery. We anticipated that the model might not retrieve a coherent answer, and that it might not be able to identify a sentence's basic constituents. The bot mislabeled in fact the verb construction as a determiner phrase, thus translating *le vistə* (I have seen him) as 'the glasses'. It probably interpreted the Apulian clitic+auxiliary compound *l.e* as the Italian article *le*, and, consequently, the past participle *vistə* as a noun, which it translated as 'glasses' (note: *viste* does not mean 'glasses' in Italian).

Table 6. Prompt: commenta: a giuannə le vistə ('comment: Giovanni, I have seen him'')

Prompt	Output (literal translation)
commenta: a giuannə le vistə comment to giovanni him seen 'comment: Giovanni, I have seen him''	*La frase "a giuannə le vistə" sembra essere in dialetto napoletano, dove "a" significa "a" e "le vistə" significa "gli occhiali". "Giuannə" potrebbe essere un nome proprio di persona. Tuttavia, dal momento che manca il contesto,*

	non è possibile capire esattamente il significato preciso della frase.
	"The phrase "I have seen Giovanni" seems to be in the Neapolitan dialect, where "to" means "to" and "I have seen him" "the glasses". "Giovanni" could be a person's proper noun. However, given the lack of context, it's not possible to exactly understand the phrase's exact meaning."

Finally, we explore a regional variety of Italian spoken in the same area. In the prompt, we used a left-peripheral internal argument (*Michele*). Roughly translated, the sentence we used means *is it (exactly) Michele that you saw?*. As shown in table 7, in this case, ChatGPT wrongly identified Michele as a vocative, *Michele, did you see?* We interpret this as an indication that the model does not contain data from regional varieties of Italian, or that at least in this case, it is biased against fronted internal arguments without an object clitic. In this instance, speakers interpret *Michele* as a direct object and not as a vocative thanks to specific intonational contours, a piece of information that is missing in the provided textual input.

Table 7. Prompt: Commenta: Michele hai visto? ('comment: is it (exactly) Michele that you saw?'')

Prompt	Output (literal translation)
Commenta: Michele hai visto? comment Michele you.have seen *'comment: is it (exactly) Michele that you saw?'*	*La frase "Michele hai visto?" è una domanda rivolta a Michele, con l'obiettivo di sapere se ha visto qualcosa o qualcuno in particolare. Dal punto di vista grammaticale, la frase è composta da un soggetto ("Michele") e un verbo ("hai visto"), con l'aggiunta dell'ausiliare "hai" per indicare il tempo passato prossimo.* 'The phrase "Michele have you seen?" is a question posed to Michele, with the aim of knowing if he has seen something or someone in particular. From the grammatical point of view, the phrase is made up by a subject ("Michele") and a verb ("you have seen"), with an added auxiliary "you have" to indicate the present perfect tense.'

This suggests that these models are not good models of human linguistic competence (see also Dentella et al. 2023), at least with respect to the syntactic structures and the languages investigated in this paper.

8

Conclusion

In this paper, we aimed to answer the following questions:

i. Does ChatGPT find the reordering archaic because these orderings are rare/absent in the training?
ii. What research questions can be raised with respect ChatGPT's metalinguistic knowledge through prompt engineering?

As for (i), we performed an observational study on general-domain corpora and we observed that reorderings are not rare nor archaic.

As for (ii), we focused on prompting different complex structures in different varieties. Our prompts are however not an exhaustive method and future studies should take into account more comprehensive methods to extract grammatical information from LLM.

Future research should also focus on how LLMs may encode grammaticality (cf. Haider 2023 on English), despite the fact that NNs can encode forms of felicitousness of the sentence (Samo & Chen 2022).

On the other hand, the investigation of the output of ChatGPT (as well, different types of conversational AI based on LLMs) as corpora could be worth investing in syntactic research, in the spirit of quantitative and computational works in detecting micro- and macro-variation (Merlo 2015 and related works; Van Cranenbroek et al. 2019; Pescarini 2021, 2022; Crisma et al. 2021 *inter alia*).

References

ALFIERI, L., TAMBURINI, F., 2016. (Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format. *CEUR WORKSHOP PROCEEDINGS*, 1749, 19–23. Torino: Accademia University Press. https://dx.doi.org/10.4000/books.aaccademia.1683

ARAVIND, A., HACKL, M., & WEXLER, K., 2018. Syntactic and pragmatic factors in children's comprehension of cleft constructions. *Language Acquisition*, 25(3), 284–314. https://doi.org/10.1080/10489223.2017.1316725

BANG, Y., CAHYAWIJAYA, S., LEE, N., DAI, W., SU, D., WILIE, B., LOVENIA, H., JI, Z., YU, T., CHUNG, W., DO, Q. V., XU, Y., FUNG, P., 2023. A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint* arXiv:2302.04023. https://doi.org/10.48550/arXiv.2302.04023

BASILE, R., 2021. Neapolitan language documentation: a transcription model. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/WR2BS

BELLETTI, A., CHESI, C., 2014. A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production. *Rivista di Grammatica Generativa*, 36, 1–28.

BENDER, E. M., GEBRU, T., MCMILLAN-MAJOR, A., SHMITCHELL, S., 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

BOSCO, C., MONTEMAGNI, S., & SIMI, M., 2013. Converting italian treebanks: Towards an italian Stanford dependency treebank. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, 61–69. The Association for Computational Linguistics. http://hdl.handle.net/2318/147938

CHAVES, R. P., RICHTER, S. N., 2021. Look at that! BERT can be easily distracted from paying attention to morphosyntax. *Proceedings of the Society for Computation in Linguistics*, 4(1), 28–38. https://doi.org/10.7275/b92s-qd21

CHOMSKY, N., 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

CIGNARELLA, A. T., BOSCO, C., PATTI, V., LAI, M., 2018. Application and analysis of a multi-layered scheme for irony on the Italian Twitter Corpus TWITTIRO. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 4204–4211. https://aclanthology.org/L18-1664.pdf

CINQUE, G., 1990. *Types of Ā-dependencies*. Cambridge, MA: MIT Press.

VAN CRAENENBROECK, J., VAN KOPPEN, M., & VAN DEN BOSCH, A., 2019. A quantitative-theoretical analysis of syntactic microvariation: Word order in dutch verb clusters. *Language*, 95(2), 333–370. https://doi.org/10.1353/lan.2019.0033

CRISMA, P., GUARDIANO, C., & LONGOBARDI, G., 2020. Syntactic diversity and language learnability. *Studi e Saggi Linguistici*, 58, 99–130. https://dx.doi.org/10.4454/ssl.v58i2.265

DENTELLA, V., MURPHY, E., MARCUS, G., LEIVADA, E., 2023. Testing AI performance on less frequent aspects of language reveals insensitivity to underlying meaning. *arXiv preprint* arXiv:2302.12313. https://doi.org/10.48550/arXiv.2302.12313

GULORDAVA, K., BOJANOWSKI, P., GRAVE, E., LINZEN, T., BARONI, M., 2018. Colorless green recurrent networks dream hierarchically. *arXiv preprint* arXiv:1803.11138. https://doi.org/10.48550/arXiv.1803.11138

HAIDER, H., 2023. *Is Chat-GPT a grammatically competent informant?* Available from: https://lingbuzz.net/lingbuzz/007285

LEDGEWAY, A., 2011. Subject licensing in CP. Mapping the Left Periphery. *The Cartography of Syntactic Structures*, 5, 257–296.

KAYNE, R. S., 1994. *The Antisymmetry of Syntax*. Vol. 25. Cambridge: MIT Press.

KATZIR, R., 2023. *Why large language models are poor theories of human linguistic cognition. A reply to Piantadosi (2023)*. Available from: https://lingbuzz.net/lingbuzz/007190

LINZEN, T., BARONI, M., 2021. Syntactic structure from deep learning. *Annual Review of Linguistics*, 7, 195–212. https://doi.org/10.1146/annurev-linguistics-032020-051035

DE MARNEFFE, M.-C., GRIONI, M., KANERVA, J., GINTER, F., 2018. Assessing the annotation consistency of the universal dependencies corpora. In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, 108–115. https://aclanthology.org/W17-6514.pdf

MASSARO, A., 2022. Romance genitives: agreement, definiteness, and phases. *Transactions of the Philological Society*, 120(1), 85–102. https://doi.org/10.1111/1467-968X.12229

MASSARO, A., 2023. Adverbial Agreement: Phi Features, Nominalizations, and Fragment Answers. *Revue Roumaine de Linguistique*, 68(4), 353–375.

MERLO, P., 1994. A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, 23, 435–457. https://doi.org/10.1007/BF02146684

MERLO, P., 2016. Quantitative computational syntax: some initial results. *IJCoL. Italian Journal of Computational Linguistics*, 2(2-1). https://doi.org/10.4000/ijcol.347

MERLO, P., ACKERMANN, F., 2018. Vectorial semantic spaces do not encode human judgments of intervention similarity. In *Proceedings of The 22nd Conference on Computational Natural Language Learning*, 392–401. http://dx.doi.org/10.18653/v1/K18-1038

MERLO, P., SAMO, G., 2022. Exploring T3 languages with quantitative computational syntax. *Theoretical Linguistics*, 48(1-2), 73–83. https://doi.org/10.1515/tl-2022-2032

NIVRE, J., 2015. Towards a Universal Grammar for Natural Language Processing. In A. Gelbukh (Ed.), *International Conference on Intelligent Text Processing and Computational Linguistics: 16th International Conference, CICLing 2015, Proceedings, Part I, (Cairo, Egypt, April 14-20, 2015)*. Cham: Springer. https://doi.org/10.1007/978-3-319-18111-0_1

OpenAI, 2023. *GPT-4 Technical Report*. Available from: https://cdn.openai.com/papers/gpt-4.pdf (accessed on March 22, 2023).

PESCARINI, D., 2022. A quantitative approach to microvariation: negative marking in central Romance. *Languages*, 7(2), Article 87. https://doi.org/10.3390/languages7020087

RADFORD, A., NARASIMHAN, K., SALIMANS, T., SUTSKEVER, I., 2018. *Improving Language Understanding by Generative Pretraining*. Available from: URL: https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

SAMO, G., BONAN, C., & SI, F., 2022. Health-Related Content in Transformer-Based Deep Neural Network Language Models: Exploring Cross-Linguistic Syntactic Bias. *Studies in health technology and informatics*, 295, 221–225. https://doi.org/10.3233/SHTI220702

SAMO, G., BONAN, C., 2023. Health-Related Content in Transformer-Based Language Models: Exploring Bias in Domain General vs. Domain Specific Training Sets. *Studies in health technology and informatics*, 302, 743–744. https://doi.org/10.3233/SHTI230252

SAMO, G., CHEN, X., 2022. Syntactic locality in Chinese in-situ and ex-situ wh-questions in transformer-based deep neural network language models. *Paper presented at Workshop on Computational Linguistics on East Asian Languages (the 29th International Conference on Head-Driven Phrase Structure Grammar), July 31st 2022, online event.*

SAMO, G., MERLO, P., 2019. Intervention effects in object relatives in english and italian: a study in quantitative computational syntax. In *Proceedings of SyntaxFest, Paris, France*, 46–56. https://aclanthology.org/W19-7906.pdf

SAMO, G., MERLO, P., 2021. Intervention effects in clefts: a study in quantitative computational syntax. *Glossa: a journal of general linguistics*, 6(1), Article 145. https://doi.org/10.16995/glossa.5742

SAMO, G., MERLO, P., 2023. Distributed computational models of intervention effects: a study on cleft structures in French. In C. Bonan & A. Ledgeway (Eds.), *It-clefts: Empirical and Theoretical Surveys and Advances* (pp. 157–180). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110734140-007

SINHA, K., JIA, R., HUPKES, D., PINEAU, J., WILLIAMS, A., KIELA, D., 2021. Masked language modeling and the distributional hypothesis: Order word matters pre-training for little. *arXiv preprint* arXiv:2104.06644. https://doi.org/10.48550/arXiv.2104.06644

SANGUINETTI, M., BOSCO, C., LAVELLI, A., MAZZEI, A., TAMBURINI, F., 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. *Proceedings of LREC 201*. https://aclanthology.org/L18-1279

ROLAND, D., O'MEARA, C., YUN, M., MAUNER, G., 2007. Processing object relative clauses: Discourse or frequency. Poster presented at the *CUNY Sentence Processing Conference. La Jolla, CA*.

WALLAT, J., SINGH, J., & ANAND, A., 2021. BERTnesia: Investigating the capture and forgetting of knowledge in BERT. *arXiv preprint* arXiv:2106.02902. https://doi.org/10.48550/arXiv.2106.02902

WILCOX, E., LEVY, R., & FUTRELL, R., 2019. Hierarchical representation in neural language models: Suppression and recovery of expectations. *arXiv preprint* arXiv:1906.04068. https://doi.org/10.48550/arXiv.1906.04068

WILCOX, E. G., FUTRELL, R., & LEVY, R., 2022. Using computational models to test syntactic learnability. *Linguistic Inquiry*, 1–88. https://doi.org/10.1162/ling_a_00491

ZEMAN, D., NIVRE, J., ABRAMS, M. M., et al., 2022. *Universal Dependencies 2.11*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Available from: http://hdl.handle.net/11234/1-4923. https://doi.org/10.1162/coli_a_00402

Angelapia Massaro: Adjunct Professor, University of Siena, Palazzo San Niccolò - via Roma, 56, 53100 Siena – Italy, angelapia.masaro@unisi.it

Giuseppe Samo: Professor, Beijing Language and Culture University, Department of Linguistics, Xue Yuan Road 15, Mailbox 82, 100083, Beijing (China), samo@blcu.edu.cn ; Research Associate, University of Geneva, Department of Linguistics, Rue de Candolle 2, 1203 Geneva (Switzerland), giuseppe.samo@unige.ch.