Syntactic patterns of Italo-Romance CP-layers in Transformers, ChatGPT and Deepseek: two case studies from Romansh and Neapolitan

Giuseppe Samo (Beijing Language and Culture University), *Angelapia Massaro* (University of Siena)[1]

samo(at)blcu.edu.cn, angelapia.massaro(at)unisi.it

Abstract

Large-language models (LLMs) have recently become the object of syntactic investigation, whether via standard probability scores in masked modeling or the interaction with conversationalAI. This study reports the results on some syntactic patterns of the Left Periphery of the clause in two Romance varieties: V2 and violations to V2 in Romansh and topic-subject agreement in Neapolitan. We study masking models with multilingual transformers (multilingual BERT and monolingual via adapters, Swiss-BERT for Romansh; multilingual BERT for Neapolitan) and through interactions with two ConversationalAI (ChatGPT, DeepSeek) by prompting an evaluation task in three languages (English, Italian for ChatGPT, Chinese for DeepSeek). Our results show asymmetries across models and structures: the theoretical predictions for Romansh are confirmed by the monolingual transformer and partially by conversationalAIs, while the Neapolitan topic-subject agreement remains challenging.

Keywords

Left periphery, cartography, ChatGPT, DeepSeek, transformers, Italian

1 AI on AI in language generation in general

Language models have demonstrated impressive abilities in linguistic tasks (Waldis et al. 2024) and have been the object of linguistic investigation, serving as "participants" for "psycholinguistic" experiments to explore their linguistic knowledge (Linzen et al. 2016; Gulordava et al. 2018; Wilcox et al. 2024 *inter alia*). Much of the investigation is based on English, but recent developments also have focused on the linguistic ability of LLMs in Romance languages like French (An et al. 2023; Nastase et al. 2024a; Samo and Merlo 2024 and references therein), Italian (cf. Attanasio et al. 2024), and Romanian (Nastase et al. 2024b). On the other hand, much less work has appeared on less-resourced Romance languages such as Italo-Romance varieties (but do see Massaro and Samo 2024). The landscape of Italo-Romance varieties is in fact characterized by substantial macro and microvariation, and these languages differ significantly from standard Italian in terms of lexicon, phonology, morphology, and syntax (Manzini and Savoia 2005; Di Caro 2022, Chapter 1).

[1] The present paper is the result of joint work. However, Giuseppe Samo and Angelapia Massaro take joint responsibility for Section 1, Giuseppe Samo for Section 2, Angelapia Massaro for Section 3.1, Giuseppe Samo for Section 3.2, Angelapia Massaro for Section 4.3, and Giuseppe Samo for Section 5.

AILing

Samo, Giuseppe & Massaro, Angelapia. Syntactic patterns of Italo-Romance CP-layers in Transformers, ChatGPT and Deepseek
Special Issue: *AI-Driven Linguistic Studies*, Vol.2. No.2
DOI: 10.62408/ai-ling.v2i2.25

AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses ISSN: 2943-0070

In this paper, we adopt a cartographic approach to generative grammar, and we focus on two phenomena related to the Left Periphery of the clause (henceforth LP, Rizzi 1997; Rizzi and Bocci 2017). The LP constitutes the highest layer of syntactic structure and hosts functional lexicon elements like complementizers, as well as elements associated with semantic and pragmatic properties, such as topicalized and focalized constituents in Romance. Specifically, we analyze violations (V3) of Verb Second (V2) orders in Eastern Romansh varieties (1), and topicalization strategies of subjects in Neapolitan (2).

(1) *Uossa Ciglia la clama*
Then Ciglia 3SG.F.CL calls
'Then, Ciglia calls her' (Putèr, Oetzel 1994: 157)

(2) *Chillo nipoteme è nu buon guaglione*
That-SM son-my-SM is a good guy
'As for my nephew, he's a good guy' (Neapolitan)

Both structures are indeed described as activating selected sections of the LP. In (1), the fronted adverbial (*uossa* 'then') targets a position higher than the usual complementizer position (SpecModP as in Samo 2022 or higher than ForceP, Wolfe 2019), while the demonstrative-like element in (2) is instead possibly generated in TopicP (Ledgeway 2011). Both phenomena also deal with two core elements of grammar: positions and dependencies. The configuration in (1) is related to the position of the verb, while the structure in (2) represents a phenomenon of morphological agreement (in terms of number and gender) between the sentence-initial demonstrative and an NP within an Inflectional Phrase. Here, we aim to study how these phenomena are handled in transformer-based large language models.

Both linearization and agreement phenomena are well-studied in NLP at the interface with linguistic theory, where the machine acts as a participant of a psycholinguistic experiment (Gulordava et al. 2018; Wilcox et al. 2024). We implement two methodologies for each phenomenon in our experiments. In both studies, the first experiment explores Masked Language Models (MLM) on a multilingual model, while the second experiment collects data from interactions with two ConversationalAI tools — ChatGPT 4o (OpenAI 2023) and DeepSeek R1 (Guo et al. 2025) —based on LLMs.[2] By comparing the behaviour of classical masking models in bidirectional encoders with conversational LLMs, our goal is not to evaluate architectural distinctions but to establish a shared empirical ground on which the two families of models can be tested in controlled settings. Importantly, the paper does not make claims about architectural differences per se.

The two studies differ in terms of the Romance variety chosen and in the quality of the data. In the first study (Study 1), we explore synthetic data adapted from natural data collected from a corpus of fairy tales (Samo 2022 based on

[2] All the interactions with ConversationalAI have been performed on 8/02/2025.

Lebrument, ed. 2005). On the other hand, we devised ex-novo sentences for Neapolitan[3] in Study 2.

To reach our goal, we proceed as follows. In Section 2 and 3, we present the theoretical background for the phenomenon under investigation and introduce the methodology and the relevant technology used for each study, as well as the results. In Section 4 we provide a discussion about the results of both studies and how they interact with theoretical considerations. Finally, Section 5 concludes.

2 On V3 orders in Eastern Swiss Romansh

2.1 The linguistic phenomenon

Romansh consists of varieties spoken in the Southeast of Switzerland (Benincà and Haiman 1992; Anderson 2005; 2016). For the pan-dialectal standard of Rumantsch Grischun see Caduff et al. 2008). Romansh varieties are broadly referred to as "strict V2" languages, in line with Benincà and Haiman's (1992: 150) assertion that these varieties "avoid V3" (cf. Kaiser and Hack 2013: 77), despite V3 orders being possible in selected contexts (Grünert 2019), and especially in Eastern varieties. We isolate Putèr, a variety that shows a very limited set of V3 orders when an adverb or a temporal PP is the sentence-initial constituent.[4] Compare the sentence in (1), here repeated, with (3), from the western variety of Sursilvan.

(1) *Uossa Ciglia la clama*
Then Ciglia 3SG.F.CL calls
'Then Ciglia calls her'
(Putèr, Oetzel 1994: 157)

(3) **Ussa el bab clama la onda*
Then the father calls the aunt
'Then, the father calls the aunt'
(Sursilvan, adapted from Samo 2019: 43; 63c)

This set of constituents comprises a reduced family of adverbials that trigger both V2 and V3 word orders, even within the same text. According to Samo (2022), V3 orders represent cases of "highlighted" but not topicalized adverbials, in line with Rizzi (2004). Specifically, these adverbials have a temporal nature, such as *uossa* ('then') and *inaquella* ('at that point'), but also include complex phrases like *fin saira* ('until dawn'), *ün bel di* ('one day'), or *tuot in üna vuot* ('suddenly'). The data discussed in Samo (2022) confirm the predictions of a criterial approach in syntactic cartography outlined in Samo (2019): in a V3 configuration, if the initial constituent is an adverbial, only the subject can be the second element of the V3 structure.

[3] As for the problem of writing conventions for the encoding, we have decided to adopt Wikipedia's forms for Neapolitan. Also, one of the authors is a native speaker of Neapolitan who controlled for the plausibility of sentences.

[4] Within the Romance landscape, V3 orders are well studied phenomenon in the Ladin varieties spoken in Veneto and South-Tyrol (cf. Casalicchio and Cognola 2018; 2020 and reference therein).

Building on the data and discussion in Samo (2022), we ask the first research question that we aim to investigate in our first study: how do LLMs evaluate V2 and V3 orders with such adverbials?

The materials and methods are presented in Sub-section 2.2.

2.2 Materials & Methods

As mentioned in Section 1, each study explores two families of methodologies. The first experiment employs masked language modeling in transformer models, examining the scores assigned to a masked element in a given configuration, using output from a deep, multi-layer bidirectional model (BERT, Devlin et al. 2019). After processing a sequence of words in a sentence, a transformer assigns probabilities to potential next/masked words. Such a score is based on training from exposure to unannotated (raw) text. Language models, tested with various architectures, have been shown to effectively capture syntactic properties (Gulordava et al. 2018; Linzen and Baroni 2021 for an overview; Wilcox et al. 2024), including elements of word order (cf. Dufter et al. 2022).

For our data, we test two multilingual models. BERT multilingual base model (google-bert/bert-base-multilingual-cased; Devlin et al. 2019) and Swiss-BERT (*ZurichNLP/swissbert*, Vamvas et al. 2023). Based on previous results showing excellent performance with Romance languages (Acs et al. 2024), we have decided to use multilingual BERT (trained on the 104 largest languages on Wikipedia). Although Romansh is missing from this list,[5] the language model includes training data from syntactically similar languages (e.g., German) and languages with similar vocabulary (e.g., Lombard). On the other hand, Swiss-BERT includes data from Romansh varieties and the standard form.

In particular, we focus on five configurations, explored with synthetic data: (i) a standard V2 order in which the first constituent is a subject (SubjV2), (ii) a V2 configuration in which an argument of the VP shell is the first constituent (ArgV2), (iii) a V3 structure in which the subject intervenes between the argument and the inflected verb (ArgV3), (iv) a V2 order in which an adverbial triggering V3 (as discussed in Samo 2022) precedes the verb (AdvV2) and, (v) a sentence in which the adverb triggers a V3 structure with a subject in second position (AdvV3). The structures are summarized in Table 1, exemplified with a sentence. We only lexicalize subjects, a transitive verb and an argument of the VP shell (either NP or PP) and an adverb triggering V3 orders.

Table 1: Examples for each structure.

Structure	**Example**
SubjV2	*Ciglia discorra Rumantsch tuot in üna vuot* '*Ciglia* speaks Romansh suddenly'
ArgV2	*Rumantsch discorra Ciglia tuot in üna vuot*
ArgV3	*Rumantsch Ciglia discorra tuot in üna vuot*
AdvV2	*Tuot in üna vuot discorra Ciglia Rumantsch*
AdvV3	*Tuot in üna vuot Ciglia discorra Rumantsch*

[5] https://github.com/google-research/bert/blob/master/multilingual.md#list-of-languages (last accessed on 23/02/2025).

As for the lexicalization of these structures, we have built our vocabulary on the basis of adverbials triggering V3 orders in Eastern varieties (Samo 2022 based on fairy tale data extracted from Lebrument, ed. 2005; a full list is given in Samo 2022: 14, fn 16) and constituents discussed in the exemplar sets in Anderson (2016).
Table 2 summarizes the materials, listing them into three main categories: (i) subjects, (ii) elements of the VP-shell (verb and a NP/PP argument) and (iii) adverbial elements triggering V3 orders. Their combination led to a total of 726 stimuli.

Table 2: Lexical materials for each constituent class.

Constituent class	**Lexical materials**
Subjects	*il bab* 'the father', *la madrastra* 'the step-mother', *Cilgia* 'Cilgia', *la vuolpetta* 'the little-fox', *il prinz* 'the prince', *meis figl* 'my son'
VP-shell	*guardet cun öglieda chi paraiva trista* 'sadly looked', *turnet a chesa* 'came back home', *mangia da bun gust* 'eat gladly', *lascha cruder el crapun* 'drop the stone', *discorra Rumantsch* 'speak Romansh'
Adverbs	*inaquella* 'at that point', *uossa* 'then', *tuot in üna vuot* 'suddenly', *fin saira* 'till dawn', *ün bel dì* 'a day'

Multilingual BERT and Swiss-BERT minimally vary in the methodology used due to vocabulary issues of the latter model encountered in pilot experiments. For the experiment with multilingual BERT, we aimed to retrieve the mean score of a sentence by evaluating the score of each constituent for each sentence and each configuration. For example, in the sentence *Ciglia discorra Rumantsch tuot in üna vuot* (see Table 1) we masked each constituent at a time: *<mask> discorra Rumantsch tuot in üna vuot*, *Ciglia <mask> Rumantsch tuot in üna vuot, Ciglia discorra <mask> tuot in üna vuot, Ciglia discorra Rumantsch <mask>*). In contrast, for Swiss-BERT, we only mask the subject and calculate the score for 3rd person pronominal entities (*el*, *ella*) in the missing gap for the subject (e.g. *<mask> discorra Rumantsch tuot in üna vuot, Rumantsch discorra <mask> tuot in üna vuot*, *Rumantsch <mask> discorra tuot in üna vuot*, *tuot in üna vuot <mask> discorra Rumantsch, tuot in üna vuot discorra <mask> Rumantsch*), for a total of 160 stimuli. It is important to remark that Swiss-BERT allows for a language adapter (in our case Romansh, "`rm_CH`").

Let us move to the interactions with conversationalAIs based on LLMs. We explore two conversationalAIs by prompting three languages: English for ChatGPT 4o, Italian for *Italiano* ChatGPT (OpenAI 2023; see a critical discussion in Bang et al. 2023), and Chinese for DeepSeek (Guo et al. 2025). Jiang et al. (2024) have shown that in zero-shot and few-shot learning the English prompt outperforms the Italian prompt for a task on Italian.

Following previous literature, we use the conversationalAI as an informant (Haider 2023 *inter alia*), limiting the request to five groups of sentences for each interaction (Samo and Ursini 2024). Specifically, in each prompt (given in Table 3) we ask the model to output a description of the sentence and rate it on a 3-point scale (one for each sentence). Each prompt contains the relevant task description (Chinese for DeepSeek, English for ChatGPT 4o or Italian for *Italiano* ChatGPT) and the sentences set to be evaluated.

Table 3: Section and text for the prompting.

Section	**Text**
Task - Chinese	我将向您展示一系列罗曼语的句子。每组有五个句子，每句包含的单词相同，但顺序不同。请您通过提供详细的描述和打分来评判这些句子的正确性，评分从1分（很差）到3分（很好）。
Task - English	I will show you a series of sets of five sentences each in a variety of Romansh with the same words, but in different orders. You could rate the sentences with (i) a description and (ii) a rating from 1 to 3 for you how correct they are.
Task - Italian	Ti mostrerò dei gruppi di cinque frasi ciascuno in una varietà del Romancio con le stesse parole, ma con ordini diversi. Potresti valutare le frasi con (i) una descrizione e (ii) una valutazione da 1 a 3 per te quanto sono corrette.
Sentence set	il bab turnet a casa ün bel dì a casa turnet il bab ün bel dì a casa il bab turnet ün bel dì ün bel dì turnet il bab a chesa ün bel il bab dì turnet a chesa la vuolpetta guardet cun öglieda chi paraiva trista inaquella cun öglieda chi paraiva trista guardet la vuolpetta inaquella cun öglieda chi paraiva trista la vuolpetta guardet inaquella inaquella guardet la vuolpetta cun öglieda chi paraiva trista inaquella la vuolpetta guardet cun öglieda chi paraiva trista la madrastra lascha cruder el crapun uossa el crapun lascha la madrastra cruder uossa el crapun la madrastra lascha cruder uossa uossa lascha la madrastra cruder el crapun uossa la madrastra lascha cruder el crapun il prinz discorra Rumantsch uossa Rumantsch discorra il prinz uossa Rumantsch il prinz discorra uossa Uossa discorra il prinz Rumantsch uossa il prinz discorra Rumantsch Cilgia mangia da bun gust fin saira Da bun gust mangia Cilgia fin saira Da bun gust Cilgia mangia fin saira Fin saira mangia Cilgia da bun gust Fin saira Cilgia mangia da bun gust

The results will be presented and discussed in Section 2.3 and organized as follows. We first present the results for transformer-based language models, and namely the mean output score for each structure for the multilingual BERT; this is followed by the results for mean scores for 3rd person singular pronominal subjects for Swiss-BERT. We then provide the score for each structure output by the ConversationalAI based on LLMs.[6]

[6] All materials and output can be found at the following link: https://github.com/samo-g/LP-Nap-Rom.git

Hypotheses:

Following Samo (2022), we expect a particular order of ranking. SubjV2, representing the canonical clause, should be the structure with the highest score/rating, followed by the non-canonical structures (ArgV2), V3 with adverbials (AdvV3), and finally the lowest score/rating, assigned to ArgV3. We can formalize this hypothesis as H_1, given below:

H_1: SubjV2 > ArgV2 = AdvV2 > AdvV3 > ArgV3

2.3 Results and Discussion

The results for the multilingual BERT are presented in Figure 1. As a reminder, our independent variable is the type of structure (SubjV2, ArgV2, ArgV3, AdvV2 and AdvV3) and our dependent variable is the mean score for each sentence given by the score of each constituent.

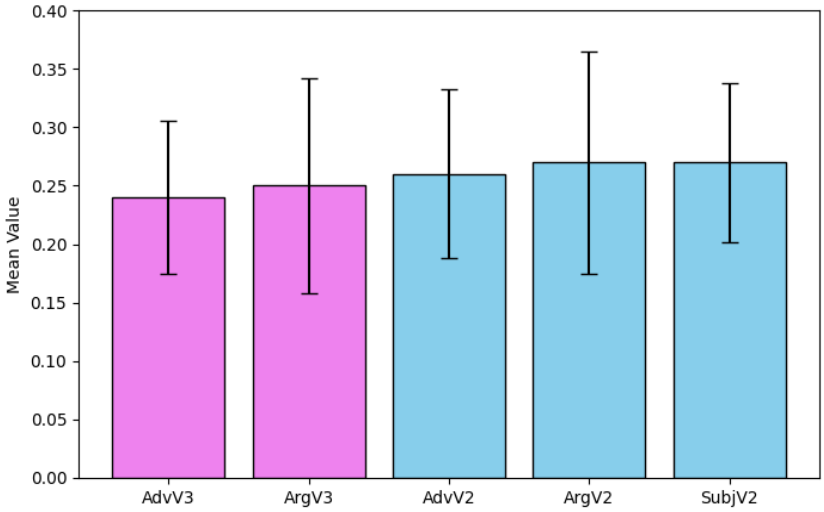

Figure 1: Multilingual BERT. Mean score for each structure. Skyblue for V2 structure, violet for V3 configurations.

SubjV2 shows the highest mean of scores (*M = 0.272, SD = 0.068*), followed by the other V2 forms and by the V3 forms, resulting in the preferred order. Our data positively shows that the V2 order is prioritized over V3 orders for both Arg and AdV3. A one-way ANOVA was conducted to compare the effect of the structure on the score across five groups. The analysis revealed a statistically significant difference among the groups, $F(4, 745) = 4.04$, $p = 0.003$, $\eta^2 = 0.0213$. Post-hoc comparisons using the Tukey HSD indicated that AdvV3 differed significantly from SubjV2 ($p = 0.0097$), with a mean difference of 0.0300 (95% CI [0.0049, 0.0552]). No other pairwise comparisons reached statistical significance.

Swiss-BERT mean scores for pronominal subjects (*el*, *ella*) in Romansh, for both V2 and V3 structures, are given in Figure 2.

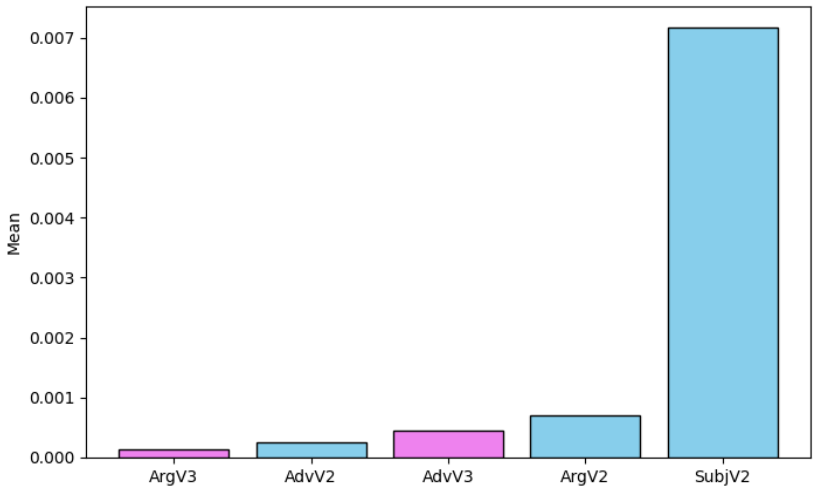

Figure 2: Swiss-BERT. Mean score for pronominal subjects. Skyblue for V2 structure, violet for V3 configurations.

Figure 2 shows a clear difference between Subject initial V2 structures and all the other marked structures. Subject initial V2 (as we have observed with multilingual models, and as we will see with conversational AIs) is the preferred configuration, marking a clear preference for what we can define as canonical clauses (Samo 2019, in line with the results on transformers on Chinese in Samo and Chen, to appear). In particular, we detect an interesting asymmetry between Arg (possibly generated in the VP-shell) and Adv (generated higher). While the V2 structure with a fronted element of the VP-shell is rated better than a V3 order, the opposite is observed for Adverbial fronting: V3 orders are rated marginally better than their V2 counterparts.

Let us now move to ConversationalAIs, whose results are summarized in Figure 3. We evaluate the quantitative output provided by the interaction, namely the rating of the sentence from 1 to 3 after a description of the sentence.[7]

[7] We leave the analysis of the qualitative content of the conversationalAI to future studies, in line with works by Mikros (2023), Samo and Ursini (2024) in standard varieties, as well as Massaro and Samo (2023) for Italian varieties.

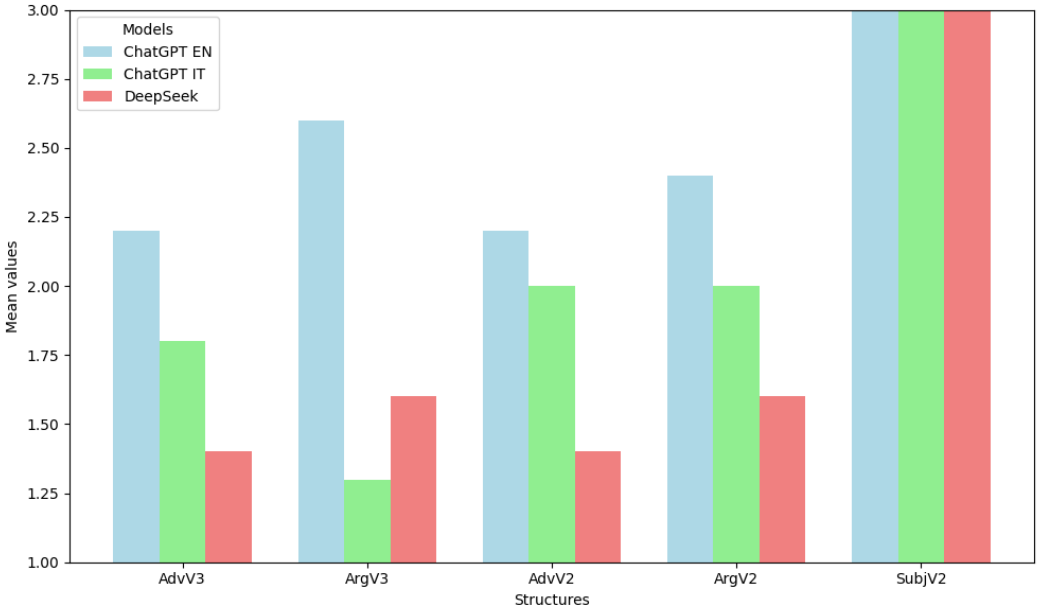

Figure 3: Mean rating (1-3) for each model.

Figure 3 shows the mean ratings (from 1 to 3) for each Romansh structure in each model. Let us start with V3 configurations with adverbials in first position. As the figure shows, ChatGPT (English) is the most permissive model, with a mean rating above 2.0. It remains the most permissive model even for other configurations, such as V3 with arguments in first position, giving a mean rating above 2.5; it gives a mean rating well above 2.0 for both V2 configurations with adverbials in first position and V2 configurations with arguments in first position. V2 configurations with subjects in first position receive instead the same rating (3 out of 3) from all models in all languages.

ChatGPT (*Italiano*) is less permissive than its counterpart in English, but still generally more permissive than DeepSeek (with the exception of V3 configurations with arguments in first position), giving a mean rating above 1.5 for V3 configurations with adverbials in first position. On average, V3 configurations with arguments in first position receive instead a mean rating below 1.5. V2 structures with adverbials and arguments in first position receive instead a mean rating of 2. As mentioned earlier, V2 structures with subjects in first position receive instead the same rating across all models (2 out of 3).

DeepSeek appears to be the least permissive model, and in line with the results from Swiss-BERT. V3 structures with adverbials in first position receive a mean rating below 1.5, while V3 configurations with arguments in first position receive an average rating slightly above 1.5. V2 structures with adverbials in first position are rated below 1.5 on average. V2 structures with arguments in first position receive instead a mean rating slightly higher than 1.5. As with the previous models, also in the case of DeepSeek V2 structures with subjects in first position are given an average rating of 3 out of 3.

Overall, we found that the output given by conversationalAIs aligns with the results from Swiss-BERT. We also found that when ChatGPT environments are in a language closer to Romansh (as in the case of *Italiano* ChatGPT) the average

rating tends to be lower than its English counterpart, the only exception being V2 structures with subjects in first position. ChatGPT 4o (English) tends in fact to be more permissive, sometimes giving a rating more than 1 point higher than its Italian counterpart (as in the case of V3 configurations with arguments in first position). Also, ChatGPT Italiano shows a light asymmetry between V2 and V3 orders.

We will revisit the key aspects of these results in Section 4. For now, we turn to the Neapolitan data.

3 On Subject topicalisation in Neapolitan

3.1 The linguistic phenomenon

Neapolitan is a southern Italo-Romance language, and, like other Romance languages, it does not have a grammatical category uniquely dedicated to the topicalization of subjects. Rather, amongst other means, it resorts to demonstrative elements in a configuration known as *double subjects* (2) (see Ledgeway 2011). As Ledgeway shows, this demonstrative has both a topic announcing and a topic shifting function.

Since Neapolitan is in a continuum with other southern Italo-Romance languages with which it shares several features (see Pellegrini 1977 for an overview), this strategy is also found in other southern Italo-Romance languages such as Gargano Apulian (Massaro 2015) or Cilentano (De Cia and Cerullo 2024). The remainder of this section, however, will be based on Ledgeway's analysis of the phenomenon in Neapolitan.

As the term *double subjects* entails, this configuration involves two elements, both superficially appearing to be subjects. In most configurations, the two elements are linearly adjacent (4a), but they can also be separated by intervening material (4b) (all following examples are adapted from Ledgeway 2011).

(4) a. **Chill-o**$_i$ **D. Michele**$_i$ pazzèa!
DEM(M)-M.SG Don Michele jokes
'Don Michele is just joking!'

b. **Chell-a**$_i$ ogge **'a** **vit-a**$_i$ s'è allungata
DEM(F)-F.SG today DEF.F.SG life-F.SG self=is lengthened
'Today life has got longer'

As shown by the indexes ($_i$/$_i$), the demonstrative and the actual subject share the same features for number and gender (gender in the root is metaphonetically realized.). This happens both when they are linearly adjacent (3a) and when they are separated by other elements, such as the temporal adverb *ogge* in (3b).[8]

[8] At least in Neapolitan, the demonstrative can also be morphologically neuter. When a neuter demonstrative is found within this construction, there is no agreement mechanism as in (3a, b). According to Ledgeway, in such instances a neuter demonstrative in a double subject construction can be interpreted as anaphoric to a proposition, and can be realised (4) as an answer to a question such as 'why is it so cold in here?': ***Chello***$_i$ *s'è astutato* ***'o riscaldamento***$_j$ DEM(N) self=is turned-off DEF.M.SG heating.M.SG'(The fact is/Because) the heating has gone off.' As the indexes ($_i$/$_j$) show,

Since neural networks and LLMs are particularly prominent in agreement phenomena (Gulordava et al. 2018; An et al. 2023), topic - subject agreement phenomena represent an excellent area of investigation. In addition, such a configuration allows for the control of possible attractors. In our case, objects of transitive verbs may represent attractors for gender agreement. Materials and methods are presented in Sub-section 3.2.

3.2 Materials & Methods

As for transformer-based models, we only operate our study on multilingual BERT. The conditions are presented in Table 4.

Table 4: Conditions and one example for each super-condition.

Conditions		**Example**
Grammatical, all elements match in gender features	MMM, FFF	*Chill o'rre sunave o'mandulino* (MMM) that.M.SG the king.M.SG played the mandolin.M.SG
Grammatical, subjects and objects mismatch in gender features	MMF, FFM	*Chill o'rre sunave a'chitarra* (MMF) that.M.SG the king.M.SG played the guitar.F.SG
Ungrammatical, the topic agrees with the object	FMF, MFM	*Chell o'rre sunave a'chitarra* (FMF) that.F.SG the king.M.SG played the guitar.F.SG
Ungrammatical, the topic does not agree with none of the arguments	FMM, MFF	*Chell o'rre sunave o'mandulino* (FMM) that.F.SG the king.M.SG played the mandolin.M.SG

Since Neapolitan does not have a standard lexicon, we have decided to work on two types of sources: frequent words in the Leipzig corpora (Goldhahn et al. 2012) of Wikipedia in Neapolitan, and two pages of Italian Wikipedia on Neapolitan paremiology and the content of a Neapolitan song.[9] The set of lexical material for each constituent family is given in Table 5, for a total of 512 stimuli.

chello, which is neuter, does not agree with *riscaldamento*, a masculine noun, in contrast to (3a, b), where the demonstrative and the subject are co-indexed. The tonic vowel in the root of the demonstrative is in fact non-metaphonetic (compare with *chillo*). Agreement-wise, this can be seen as the demonstrative agreeing with a nominal factive operator (*the fact*), which, as Ledgeway suggests, is anaphoric to a proposition. Had a neuter form not existed in Neapolitan, agreement would probably have resulted in a masculine form, as it happens in the nominalisation of events in Italian or, within southern Italo-Romance, in Gargano Apulian (see Massaro 2023).

[9] https://wortschatz.uni-leipzig.de/en/download/Neapolitan (last accessed on 26/01/2025).
https://it.wikiquote.org/wiki/Modi_di_dire_napoletani(last accessed on 26/01/2025).
https://it.wikipedia.org/wiki/Dduje_Paravise_(brano_musicale)(last accessed on 26/01/2025).

Table 5: Lexical materials for each constituent class.

Constituent class	**Lexical materials**
Subjects	*o' padrone* 'the landlord.M', *o'rre* 'the king.M', *o' segnore* 'the gentleman.M', *nu fravecature* 'the worker.M', *a' padrona* 'a landlady.F', *a'reggina* 'the queen.F', *a' segnora* 'the lady.F', *na creature* 'a child.F'
Verb + object	*sunave o'violine, na' canzone* 'played the violin.M, a song.F' *acconcia o'mandulline, na' cammisa* 'fixes the mandolin.M, a shirt.F' *s'accattaje l'organette, na cucozza* 'bought the accordeon.M, a pumpkin.F' *scassaje nu' piezz, 'a chitarra* 'broke a piece.M, the guitar.F'

Finally, the prompt partially shares some of the properties with the one presented for Romansh (cf. Table 3). We however introduce the target language, Neapolitan, and a different subdivision of sentence groups. The prompt and the sentences are given in Table 6.

Table 6: Section and text for the prompting.

Section	**Text**
Task - Chinese	我将向您展示一系列那不勒斯语的句子。每组有四个句子，每句包含的单词相同，但顺序不同。请您通过提供详细的描述和打分来评判这些句子的正确性，评分从1分（很差）到3分（很好）。
Task - English	I will show you a series of sets of four sentences each in Neapolitan with the same words, but in different orders. You could rate the sentences with (i) a description and (ii) a rating from 1 to 3 for you how correct they are.
Task - Italian	Ti mostrerò dei gruppi di quattro frasi ciascuno in napoletano con le stesse parole, ma con ordini diversi. Potresti valutare le frasi con (i) una descrizione e (ii) una valutazione da 1 a 3 per te quanto sono corrette.
Sentence set	Chill o'padrone sunave o'violine Chell o'padrone sunave o'violine Chill a'padrona sunave o'violine Chell a'padrona sunave o'violine Chill o'rre sunava na canzone Chell o'rre sunava na canzone Chill a' reggina sunava na canzone Chell a'reggina sunava na canzone Chill o'signore acconcia na cammisa Chell o'signore acconcia na cammisa Chill a' signora acconcia na cammisa Chell a' signora acconcia na cammisa chill nu fravecature s'accattaje l'organette chell nu fravecatore s'accattaje l'organette chill nu fravecatore s'accattaje na cucozza chell nu fravecatore s'accattaje na cucozza chill a creatura scassaje nu piezz chell a creature scassaje nu piezz chill a creature scassaje a chitarra chella a creature scassaje a chitarra

Hypotheses

We expect that grammatical configurations, i.e., agreement between the topical functional element (*chill*, *chell*) and the subject, will be realized as grammatical. Two possible hypotheses can be formulated: (i) one hypothesis suggests that the absence of distractors—where the object shares the same gender feature as the topic or the subject—has an ameliorative effect, or (ii) mismatched gender configurations between the subject and the object might have an ameliorative effect because the attention mechanism is more sensitive to the difference. We summarize these two hypotheses as H_{2a} and H_{2b}.

H_{2a}: MMM/FFF > MMF/FFM > FMM/MFF = MFF/MFM

H_{2b}: > MMF/FFM > MMM/FFF > FMM/MFF = MFF/MFM

Results are presented in Section 3.3.

4.3 Results and Discussion

Figure 4 below shows the mean value for ratings of sentences with a mismatch in the morphological gender of subjects and direct objects.

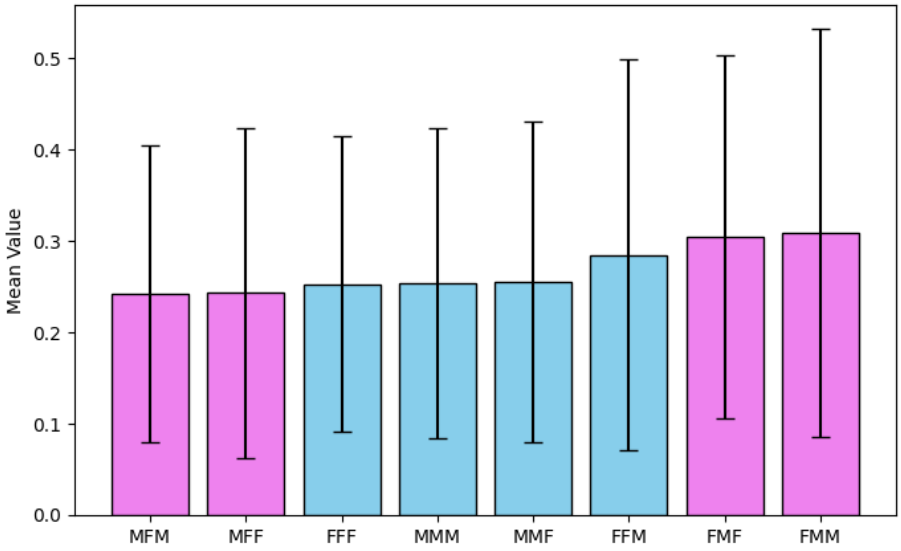

Figure 4. Mean rating of sentences and standard deviation with gender mismatches between subjects and direct objects in Neapolitan.

As Figure 1 shows, none of the hypotheses in Section 3.2. properly predict the results. As we can see, the highest mean rating is given to ungrammatical sentences where there is no mismatch between the subject and the direct object (FMM), even if the double subject (the demonstrative) is feminine instead of the expected gender (masculine - as outlined in Section 3, the two should instead agree). Mismatching between the subject (M) and the direct object (F) is instead given a slightly lower mean rating (FMF), followed by a sentence where the object (M) mismatches with

the subject (F), which instead agrees with the double subject (F) (FFM). In between we find instances where all three elements are of the same gender (FFF, MMM), and an instance where the double subjects are both masculine (MMF). Interestingly, in this case the rating is lower than cases in which the double subjects are of the same gender, but they are feminine (FFM). The lowest ratings are given to cases in which the demonstrative and the direct object have mismatching genders, but with *chill* masculine as the demonstrative (MFM, MFF). A one-way ANOVA analysis did not reveal a statistically significant difference among the groups, $F(7, 504) = 1.33$, $p = 0.234$, $\eta^2 = 0.018$.

We can safely say that this type of agreement is not learnt, at least in a multilingual model trained with data from languages similar to Neapolitan. We face the same issues discussed in section 2 with the results from the multilingual BERT in Romansh. We should add, however, that we are not aware of a language model fully trained on Neapolitan.

As for ConversationalAIs based on LLMs, the results are summarized in Figure 5.

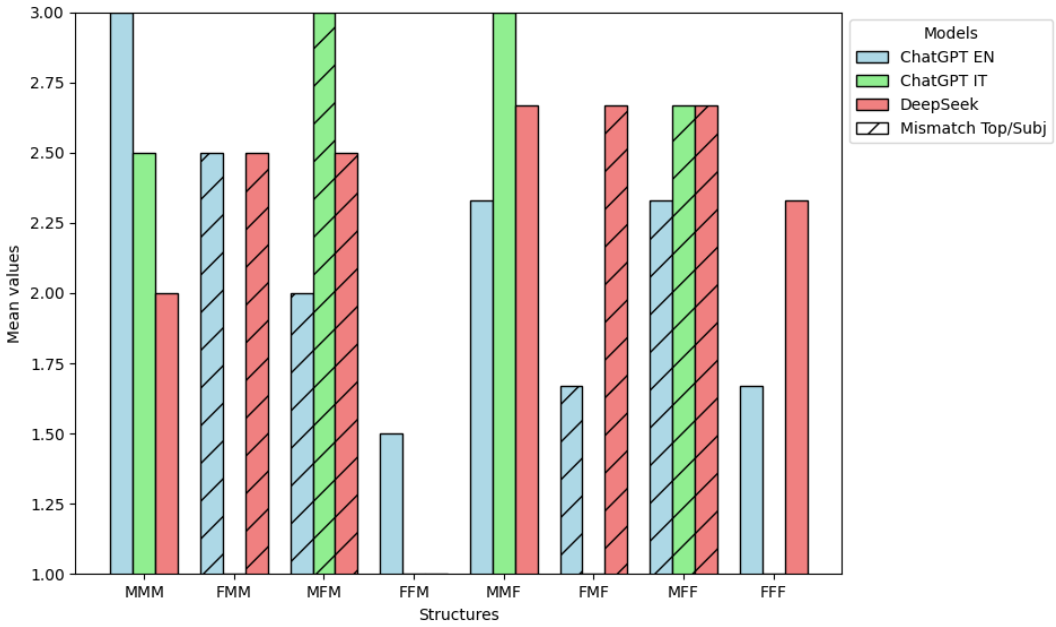

Figure 5: Mean rating (1-3) for each model.

Figure 5 shows values for Neapolitan, revealing what follows. Also in this case, none of the hypotheses predict our results. DeepSeek performs best – in terms of grammaticality of the sentences – (above 2.5) when there is a mismatch in the gender features of the topic marker (the demonstrative) and the subject (MFF, FMF, MFM, and FMM). In a single instance we find a value above 2.5 (same value as the highest) where no mismatch is present (MMF). The lowest values are assigned when there is no mismatch (FFM), but both the demonstrative and the subject are feminine (unlike MMF, whose value is instead 2.5). Feminine double subjects thus seem to trigger lower values than masculine double subjects.

ChatGPT *Italiano* performs best (value: 3) in two cases. In the first, there is a mismatch between the gender of the demonstrative and the gender of the subject

(MFM). In the second, there is no mismatch between the gender of the demonstrative and the gender of the subject (MMF). The last case also applied to DeepSeek, as we said. The lowest values are found when either the demonstrative or both the demonstrative and the subject are feminine (FMM, FFM, FMF, FFF). This last finding patterns with what also found with DeepSeek. Masculine demonstratives are thus assigned a value of 2.5 or higher, in contrast to feminine ones.

ChatGPT (English) is characterized by the following pattern. The highest value (3) is found when the demonstrative is masculine (but no mismatch between the demonstrative and the subject is present, MMM). The second-highest value (2.5) is found when the demonstrative is feminine, and there is a mismatch between the gender of the demonstrative and the gender of the subject (FMM). This is followed by values above 2 with or without a mismatch (MMF and MFF), both characterized by a masculine demonstrative in first position. Values slightly above 1.5 are found in instances where the demonstrative is feminine, regardless of whether the mismatch is present or not (FMF, FFF). This is followed by the lowest value (1.5), assigned to instances with no mismatch, but with a feminine demonstrative in first position (FFM).

In a nutshell, in all Conversational AIs and conditions considered, the lowest value is found when the first element (in our case a demonstrative topic marker) is feminine. This can be accompanied by the presence or absence of a gender mismatch, depending on the Conversational AI and the language considered. Still, a feminine element in first position is correlated with a lower value.

5 Discussion

In this paper, we have tested complex phenomena involving the LPin Romansh and Neapolitan by retrieving and analyzing data extracted from transformer-based language models and through interactions with conversational AIs based on LLMs.

The results for transformer-based models align with the current literature (see also García and Wanner 2025; Hamerlik 2025: Chapter 4; Nastase et al. 2024). Monolingual data (or at least a multilingual model containing an adapter for monolingual data, as in the case of Swiss-BERT) follows the expected trends. Multilingual models with closely related languages remain an excellent tool for low-resource languages. However, according to our investigation, they may not be as effective for complex structures involving LP elements. In this sense, we acknowledge that the Left Periphery adds an additional layer of complexity. One of the examined phenomena involves positional constraints, while the other concerns agreement.

The comparison of the quantitative output of Conversational AIs is informative. The results for Romansh are partially in line with those from the monolingual model. DeepSeek demonstrates excellent performance, as does ChatGPT *Italiano*. The methodology of comparing prompts across languages is not new in the literature (Merlo 2023 and references therein); however, this study extends the approach by incorporating prompting in Chinese for DeepSeek, in addition to Italian and English, to examine the grammatical properties of the output.

Our study contributes to the development of a novel methodology for analyzing specific phenomena in generative grammar and syntactic cartography (but not limited too): it engages with the discussion on testing theoretical frameworks (Study 1) in the spirit of a computational cartography approach (Samo, to appear) while also offering insights into agreement phenomena in language models (Study 2).

To summarize, we emphasize the importance of monolingual models for under-resourced languages and leave qualitative aspects of the output to future studies. Further research should expand the range of phenomena and languages analyzed, ideally within the Italo-Romance landscape, using a similar methodology to enable more precise comparisons. In this regard, the availability of dedicated resources—such as Savoia et al. (2025), which includes digitized content extracted from Manzini and Savoia (2005)—is a crucial factor for constructing curated datasets.

6 Conclusion

In this paper, we provide new data on complex phenomena involving the LP of the clause in two Romance varieties, Romansh and Neapolitan, extracted from monolingual (via adapter) transformer-based models and through interactions with Conversational AI (ChatGPT, with English and Italian prompting, and DeepSeek, with Chinese prompting).

These phenomena focus on position and agreement, and we observe asymmetries across models and structures. Future research should expand the number of languages and phenomena analyzed, particularly within the Italo-Romance landscape, to better predict and understand linguistic variability.

Acknowledgments

We acknowledge the use of OpenAI's GPT-4-Turbo in assisting with the editing and refinement of this paper, enhancing both its clarity and presentation in English. However, we ensured that all contributions adhered strictly to the standards and ethical guidelines of academic writing.

Funding

Giuseppe Samo acknowledges the support of the project "A comparative study of syntactic processing in Humans and LLMs" supported by the Ministry of Education Humanities and Social Sciences Post-Funding Project (24JHQ046).

Angelapia Massaro's research has been supported by Project RESYNC European Union – Next-GenerationEU – National Recovery and Resilience PLAN (PNRR) – Mission 4, Component 2, Investment 1.1 Fondo per il Programma Nazionale di Ricerca e Progetti di Rilevante Interesse Nazionale (PRIN) – CUP B53D23030300001.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this contribution.

References

Acs, Judit & Hamerlik, Endre & Schwartz, Roy & Smith, Noah A. & Kornai, Andras 2024. Morphosyntactic Probing of multilingual BERT Models. *Natural Language Engineering* 30. 753–92. http://dx.doi.org/10.1017/S1351324923000190

An, Aixiu & Jiang, Chunyang & Rodriguez, Maria A. & Nastase, Vivi & Merlo, Paola. 2023. BLM-AgrF: A new French benchmark to investigate generalization of Agreement in Neural Networks. In Vlachos, Andreas & Augenstein, Isabelle (eds), *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (Dubrovnik, May 2-6, 2023), 1363–1374. Association for Computational Linguistics. http://doi.org/10.18653/v1/2023.eacl-main.99

Attanasio, Giuseppe & Basile, Pierpaolo & Borazio, Federico & Croce, Danilo & Francis, Maria & Gili, Jacopo & Musacchio, Elio & Nissim, Malvina & Patti, Viviana & Rinaldi, Matteo & Scalena, Daniel. 2024. CALAMITA: Challenge the Abilities of LAnguage Models in ITAlian. In Dell'Orletta, Felice & Lenci, Alessandro & Montemagni, Simonetta & Sprungnoli, Rachele (eds), *Proceedings of the 10th Italian Conference on Computational Linguistics CLiC-it 2024* (Pisa, December 4-6, 2024). CEUR Workshop Proceedings. https://aclanthology.org/2024.clicit-1.116/

Anderson, Stephen R. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199279906.001.0001

Anderson, Stephen R. 2016. Romansh (Rumantsch). In Ledgeway, Adam & Maiden, Martin (eds), *The Oxford Guide to the Romance Languages*, 169–184. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0012

Bang, Yejin & Cahyawijaya, Samuel & Lee, Nayeon & Dai, Wenliang & Su, Dan & Wilie, Bryan & Lovenia, Holy & Ji, Ziwei & Yu, Tiezheng & Chung, Willy et al. 2023. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. *arXiv*. https://doi.org/10.48550/arXiv.2302.04023

Benincà, Paola & Haiman, John. 1992. *The Rhaeto-romance Languages*. London: Routledge. https://doi.org/10.4324/9780203992487

Caduff, Renzo & Caprez, Uorschla N. & Darms, Georges. 2008. *Grammatica per l'instrucziun dal rumantsch grischun*. Departament da linguatgs e litteraturas

romanas, Linguatg e litteratura rumantscha, University of Fribourg. https://folia.unifr.ch/rerodoc/9712/files/Gramminstr.pdf (last accessed on 2/02/2025).

Casalicchio, Jan & Cognola, Federica. 2018. Verb-second and (micro) variation in two Rhaeto-romance varieties of Northern Italy. In D'Alessandro, Roberta & Pescarini, Diego (eds), *Advances in Italian Dialectology*, 72–105. Leiden: Brill. https://doi.org/10.1163/9789004354395_005

Casalicchio, Jan & Cognola, Federica. 2020. Parametrising 'lexical subject-finite verb' inversion across V2 languages. On the role of Relativised Minimality at the vP edge. In Woods, Rebecca & Wolfe, Sam (eds), *Rethinking Verb Second*, 594–622. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198844303.003.0025

De Cia, Simone & Cerullo, Mariangela. 2024. Towards a typology of Zero Aboutness: expletive A in Fornese and Chiru in Cilentano. In Ojea, Ana (ed.), *Syntax and Discourse at the Crossroads, Languages* 9(2), 1–24. https://doi.org/10.3390/languages9020060

Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In Burstein, Jill & Doran, Christy & Solorio, Thamar (eds), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human Language technologies*, vol. 1 (June 2-7, 2019), 4171–4186. Minneapolis: Association for Computational Linguistics. http://doi.org/10.18653/v1/N19-1423

Di Caro, Vincenzo Nicolò. 2022. Il bilinguismo italiano-dialetto in Sicilia. Profilo sociolinguistico, nuove realtà comunicative e prospettive didattiche. Venezia: Edizioni Ca' Foscari. http://doi.org/10.30687/978-88-6969-649-7

Dufter, Philipp & Schmitt, Martin & Schütze, Hinrich 2022. Position Information in Transformers: an overview. *Computational Linguistics* 48(3). 733–763. http://doi.org/10.1162/coli_a_00445

García, Alba Táboas & Wanner, Leo. 2025. Assessing the Agreement competence of Large Language Models. In Hajičová, Eva & Kahane, Sylvain (eds), *Proceedings of the Eighth International Conference on Dependency Linguistics* (Ljubljana, August 27-28, 2025), 36–53. Association for Computational Linguistics. https://aclanthology.org/2025.depling-1.4/ (last accessed on 2/02/2025).

Goldhahn, Dirk & Eckart, Thomas & Quasthoff, Uwe. 2012. Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. In Calzolari, Nicoletta & Choukri, Khalid & Declerck, Thierry & Ugur Dogan, Mehmet & Maegaard, Bente & Mariani, Joseph & Moreno, Asuncion & Odijik, Jan & Piperidis, Stelios (eds), *Proceedings of the Eighth International Conference on Language Resources and Evaluation LREC'12*

LREC (Istanbul, May 2012), Vol. 29, 31–43. European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2012/pdf/327_Paper.pdf (last accessed on 2/02/2025).

Gulordava, Kristina & Bojanowski, Piotr & Grave, Edouard & Linzen, Tal & Baroni, Marco. 2018. Colorless Green Recurrent Networks Dream Hierarchically. In Walker, Marilyn & Ji, Heng & Stent, Amanda (eds), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1 (New Orleans, June 1-6, 2018), 1195–1205. Association for Computational Linguistics. http://doi.org/10.18653/v1/N18-1108

Guo, Daya & Yang, Dejian & Zhang, Haowei & Song, Junxiao & Zhang, Ruoyu & Xu, Runxin & Zhu, Qihao et al. 2025. Deepseek-r1: incentivizing reasoning capability in LLMs via reinforcement learning. *arXiv*. https://doi.org/10.48550/arXiv.2501.12948

Grünert, Mathias. 2019. Die Konstituentenfolg im bündnerromanischen Nebensatz. *Ladinia* XLII (2018), 23–39.

Haider, Hubert. 2023. Is Chat-GPT a grammatically competent informant? Available from: https://lingbuzz.net/lingbuzz/007285 (last accessed on 2/02/2025).

Hamerlik, Endre. 2025. *Internal Representations in Neural Network*. Unpublished PhD Dissertation. Comenius University Bratislava.

Jiang, Chunyang & Samo, Giuseppe & Nastase, Vivi & Merlo, Paola. 2024. BLM-It - Blackbird Language Matrices for Italian: A CALAMITA Challenge. In Dell'Orletta, Felice & Lenci, Alessandro & Montemagni, Simonetta & Sprugnoli, Rachele (eds), *Proceedings of the 10th Italian Conference on Computational Linguistics CLiC-it 2024* (Pisa, December 4-6, 2024). CEUR Workshop Proceedings. https://aclanthology.org/2024.clicit-1.125/ (last accessed on 2/02/2025).

Kaiser, Georg A. & Hack, Franziska M. 2013. Language change in comparison: the (special) case of Raeto-Romance. In Fleischer, Jürg & Horst, Simon J. (eds), *Sprachwandelvergleich-Comparing Diachronies*, 44–73. Berlin: De Gruyter. https://doi.org/10.1515/9783110313468.73

Lebrument, Hanspeter (ed.). 2006. *Paraulas Märchen in Sursilvan, Sutsilvan, Surmiran, Puter, Vallader und Deutsch*. Chur/Glarus/Zürich: La Quotidiana Südostschweiz Buchverlag.

Ledgeway, Adam. 2011. Subject licensing in CP: the Neapolitan Double-Subject Construction. In Benincà, Paola & Munaro, Nicola (eds), *Mapping the Left Periphery: The Cartography of Syntactic Structures*, vol. 5, 257–296. https://doi.org/10.1093/acprof:oso/9780199740376.003.0009

Linzen, Tal & Dupoux, Emmanuel & Goldberg, Yoav. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics* 4. 521–535. http://doi.org/10.1162/tacl_a_00115

Linzen, Tal & Baroni, Marco. 2021. Syntactic structure from deep learning. *Annual Review of Linguistics* 7. 195–212. https://doi.org/10.1146/annurevlinguistics-032020-051035

Manzini, M. Rita & Savoia, Leonardo M. 2005. *I dialetti italiani e romanci: Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.

Massaro, Angelapia. 2015. Il DP nel dialetto sammarchese. Master Thesis. Università degli Studi Roma Tre.

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

Massaro, Angelapia & Samo, Giuseppe. 2023. Prompting Metalinguistic Awareness in Large Language Models: ChatGPT and Bias effects on the Grammar of Italian and Italian varieties. *Verbum* 14. https://doi.org/10.15388/Verb.42

Merlo, Paola. 2023. Blackbird language matrices (BLM), a new task for rule-like generalization in neural networks: can Large Language Models pass the test? In Bouamor, Houda & Pino, Juan & Bali, Kalika (eds), *Findings of the Association for Computational Linguistics EMNLP 2023* (Singapore, December 6-10, 2023), 8119–8152. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.546

Mikros, George. 2023. AI-Writing detection using an ensemble of Transformers and Stylometric features. Talk held at Qualico 2023, University of Lausanne, 29/06/2023.

Nastase, Vivi & Samo, Giuseppe & Jiang, Chunyang & Merlo, Paola. 2024a. Exploring Italian sentence embeddings properties through multi-tasking. In Dell'Orletta, Felice & Lenci, Alessandro & Montemagni, Simonetta & Sprugnoli, Rachele (eds), *Proceedings of the 10th Italian Conference on Computational Linguistics CLiC-it 2024* (Pisa, December 4-6, 2024). CEUR Workshop Proceedings. https://aclanthology.org/2024.clicit-1.70/ (last accessed on 2/02/2025).

Nastase, Vivi & Samo, Giuseppe & Jiang, Chunyang & Merlo, Paola. 2024b. Exploring syntactic information in sentence embeddings through multilingual subject-verb agreement. In Dell'Orletta, Felice & Lenci, Alessandro & Montemagni, Simonetta & Sprugnoli, Rachele (eds), *Proceedings of the 10th Italian Conference on Computational Linguistics CLiC-it 2024* (Pisa, December 4-6, 2024). CEUR Workshop Proceedings. https://aclanthology.org/2024.clicit-1.71/ (last accessed on 2/02/2025).

Oetzel, Annette. 1994. Die Nicht-Einhaltung der Inversion im Engadinischen und ihr Einfluss auf die Informationsstruktur. *Annalas de la Societad Retorumanscha* 107, 153–171.

OpenAI. 2023. GPT-4 Technical Report. Available from: https://cdn.openai.com/papers/gpt-4.pdf (last accessed on 28/02/2025).

Pellegrini, Giovan Battista. 1977. *Carta dei dialetti d'Italia*. Firenze: Litografia Artistica Cartografica.

Rizzi, Luigi. 1997. *The Fine Structure of the Left Periphery.* In Haegeman, Liliane (ed.), 281–337. Dordrecht: Kluwer Academic publisher.

Rizzi, Luigi & Bocci, Giuliano. 2017.The Left Periphery of the clause: primarily illustrated for Italian. In Everaert, Martin & van Riemsijk, Henk C. (eds), *The Blackwell Companion to Syntax*, 1–30. Hoboken: John Wiley and Sons. https://doi.org/10.1002/9781118358733.wbsyncom104

Rizzi, Luigi. 2004. Locality and Left Periphery. In Belletti, Adriana (ed.), *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3, 223–251. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780195171976.003.0008

Samo, Giuseppe. 2019. *A Criterial Approach to the Cartography of V2*. Amsterdam/Philadelphia: John Benjamins Publishing. https://doi.org/10.1075/la.257

Samo, Giuseppe. 2022. Criterial V2: ModP as a locus of microvariation in Swiss Romansh varieties. *Probus* (34). 143–170. https://doi.org/10.1515/probus-2022-0007

Samo, Giuseppe & Chen, Xu (to appear). When Cartography meets Transformers: a study on Syntactic Locality and the functional lexicon in Chinese wh-questions. In Fuzhen, Si & Rizzi, Luigi (eds), *Cartographic Representations at the Interfaces*. Oxford: Oxford University Press.

Samo, Giuseppe & Merlo, Paola. 2024. Distributed computational models of intervention effects: a study on cleft structures in French. In Ledgeway, Adam & Bonan, Caterina (eds), *It-Clefts: Empirical and Theoretical Surveys and Advances*, 157–180. Berlin/Boston: De Gruyter. https://doi.org/10.1515/9783110734140-007

Samo, Giuseppe & Ursini, Francesco-Alessio. 2024. The Semantics of generic terms in toponyms: Formal Semantics meets multi-source Big Data. In Hu, Xuke & Purves, Ross & Moncla, Ludovic & Kersten, Jens & Stock, Kirstin (eds), *Proceedings of the second International Workshop on Geographic Information Extraction from Texts at ECIR 2024 GeoExT 2024*, (Glasgow, March 24), 1–12.

Savoia, Leonardo M. & Manzini, M. Rita & Mazzaggio, Greta & Luca Andrea, Ludovico & Vena, Mael V. & Zoli, Carlo & Baldi, Benedetta & Franco, Ludovico & Binazzi, Neri. 2025. *The Manzini & Savoia (2005) Corpus:*

Morphosyntactic Variation in Italian and Romansh Dialects. Data set. Zenodo. https://doi.org/10.5281/zenodo.14803999

Vamvas, Jannis & Graën, Johannes & Sennrich, Rico. 2023. SwissBERT: The Multilingual Language Model for Switzerland. In Ghorbel, Hatem & Sokhn, Maria & Cieliebak, Mark & Hürlimann, Manuela & de Salis, Emmanuel & Guerne, Jonathan (eds), *Proceedings of the 8th edition of the Swiss Text Analytics Conference* (Neuchatel, June 12-14, 2023), 54–69. Association for Computational Linguistics. https://aclanthology.org/2023.swisstext-1.6/ (last accessed on 2/02/2025).

Waldis, Andreas & Perlitz, Yotam & Choshen, Leshem & Hou, Yufang & Gurevych, Iryna. 2024. Holmes. A benchmark to assess the linguistic competence of Language Models. *Transactions of the Association for Computational Linguistics* 12, 1616–1647. https://doi.org/10.1162/tacl_a_00718

Wilcox, Ethan Gotlieb & Futrell, Richard & Levy, Roger. 2024. Using computational models to test Syntactic Learnability. *Linguistic Inquiry* 55(4). 805–848. https://doi.org/10.1162/ling_a_00491

Wolfe, Sam. 2019. Redefining the typology of V2 languages. The view from Medieval Romance and beyond. *Linguistic Variation* 19. 16–46. https://doi.org/10.1075/lv.15026.wol