Tracing English interference in AI-generated German. An analysis of word order and syntactic fronting

Paolo Valentinelli (Università di Trento)*

paolo.valentinelli(at)unitn.it

Abstract

Large language models (LLMs) constitute a transformative advancement in natural language processing, yet their development remains disproportionately skewed toward English. Despite the global linguistic landscape, non-English languages – including major languages like Spanish, French or Chinese – are effectively treated as *low-resource* in current LLM training paradigms. This study analyses two linguistic traits of AI-generated texts which mimic human-authored German newspaper articles and compares them with a purpose-built corpus of *real* journalistic texts. These features are (i) word order and (ii) pre-field occupation. Through quantitative and qualitative analyses of the outputs of four distinct LLMs, three key phenomena in AI's German outputs were identified: (i) a marked preference for SVO word order; (ii) reduced syntactic variability compared to human-authored texts; and (iii) the emergence of stylistically marked constructions which mirror English linear progression rather than native German sentence bracketing. While some models approximate *human-like* syntactic patterns for certain variables, this equivalence remains limited and context-dependent, which may suggest a cross-linguistic interference from the overwhelming English predominance in LLM training data. The study emphasises the linguistic implications of LLM architectures and calls attention to the urgent need for more equitable representation of world languages in natural language processing development.

Keywords

artificial intelligence, AI-generated language, ChatGPT, German syntax, newspaper language

1 Introduction

The release of ChatGPT by OpenAI in late 2022 marked a pivotal moment in the field of natural language processing. Its capacity to comprehend, generate, and contextualise linguistic input constitutes not merely a technical advancement, but a paradigmatic shift in the field, fundamentally transforming the processes by which texts are produced, interpreted, and mediated across languages and contexts. To generate this form of *human-like* language, large language models (LLMs henceforth) undergo a process of *pre-training*, during which they are exposed to an extensive amount of textual data that allows structures, meanings and uses of language to be acquired in a statistical and predictive way.[1]

* I would like to express my sincere gratitude to the two anonymous referees for their time and detailed feedback. Their constructive comments and insightful suggestions were invaluable in shaping the final version of this manuscript and significantly improved the clarity and depth of the analysis.

[1] This textual data comprises vast amounts of publicly available web content such as blog posts, forums, and news sites; books, particularly those in the public domain; and is supplemented by

AILing

Valentinelli, Paolo. 2026.
Tracing English interference in AI-generated German.
AI-Linguistica 2026, Vol. 5 No.1
DOI: 10.62408/ai-ling.v5i1.36

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

According to Brown et al. (2020), approximately 93% of the data used to train GPT-3, a foundational model by OpenAI, is in English.[2] This implies that other widely spoken languages are significantly underrepresented in this training corpus. For example, French (312 million speakers globally) accounts for only 1.8% of the data, German (134 million) 1.5%, Spanish (558 million) 0.8%, and Italian (66 million) 0.6% (Johnson et al. 2022 for the linguistic distribution in GPT-3; Eberhard et al. 2025 for global speakers amounts). As probabilistic systems that generate output token by token based on their training data, LLMs are naturally influenced by the disproportionate volume of English in their training corpora (cf. Backus et al. 2023: 303–304).

While it is true that *state-of-the-art* models often perform well enough on multilingual benchmarks (cf. Shi et al. 2022; Grattafiori et al. 2024), more granular analyses reveal significant English interference across various linguistic levels.[3] A growing body of recent studies has begun to document these distinct English *fingerprints* in other (mostly 'high-resource') languages, such as Italian.[4] However, a research gap appears to exist concerning the German language, particularly when focusing on syntax. One exception is a pilot study by Matthies (2025), which investigates reference management and agent realisation in AI-generated academic writing.

Describing the language of LLMs in absolute terms is not only a gargantuan task but arguably an impossible one (just as it would be to definitively describe human language). Instead, to assess how English interferes with AI-generated German, it is both necessary and practical to limit the analysis to specific linguistic features within a well-defined text type. In selecting a comparative baseline for this study, we therefore turn to the journalistic prose of national newspapers.

Schmitt (2004: 20) characterises 'newspaper language' as a "Zwischenstufe zwischen Fachsprache und Umgangssprache" ['intermediate stage between specialised language and colloquial language'; P.V.], however, this term is

proprietary datasets obtained through partnerships that grant access to non-public content, such as language corpora and multimedia archives (cf. OpenAI 2023).

[2] Data concerning the language distribution within the training sets of LLMs is outdated. This issue arises primarily because most companies have ceased public disclosure of this information, leading to a deficit of transparency regarding the composition of their proprietary training data.

[3] While the term 'interference' traditionally denotes cognitive inhibition in human second language acquisition, in this study we operationalise it within the framework of statistical language modeling. We define English interference not as a cognitive lapse, but as a specific instance of negative cross-lingual transfer. Given the overwhelming prevalence of English in the pre-training data, the model's probability distribution for the target language is heavily conditioned by English structural priors. Consequently, 'interference' here refers to structural convergence, where the model generates target-language output that maps isomorphically onto English syntactic templates, disregarding the typological constraints of the target language.

[4] This trend has been confirmed by several studies focusing on Italian (e.g., Cicero 2023, 2025; De Cesare 2023a, 2023b, 2024, 2025; Tavosanis 2024). Furthermore, Guo et al. (2025) investigate French and Chinese, while Qin et al. (2025) provide a multilingual study concerning the accuracy of LLM outputs.

inherently complex in the literature, exhibiting substantial variation based on factors such as thematic focus, text type, text function, and the author's style.[5] At the same time, as a prototypical realisation of the 'language of distance' (Koch and Oesterreicher 1985), a newspaper text is characterised by a high degree of planning, syntactic complexity, and adherence to normative grammar. Thus, it is possible to observe distinct trends and similarities in regard to their syntax. This study focuses on two such features, which are found in the literature and are also expected to be influenced by English interference. These are (i) word order and (ii) pre-field occupation.

Precisely because this register is conservative and resistant to fleeting trends, it serves as the ideal *gold standard* for evaluating LLMs. Since models are typically fine-tuned with Reinforcement Learning from Human Feedback (RLHF) to act as helpful, formal assistants, their output targets this same high-register, standardised code. Therefore, any structural deviation observed in the AI output – specifically regarding English-like linearisation – cannot be dismissed as informal stylistic variation but must be analysed as a potential failure of the model to replicate the target syntax.

Lastly, newspaper language is socially pervasive, and, perhaps most critically, an increasingly relevant target for automation by AI (cf. Newman et al. 2023, 2024, 2025). As such, this paper aims to participate in the lively discourse about linguistic consequences of current training data (mis-)representation in LLMs. The findings presented in this paper strongly suggest that the predominance of English in training data leads to a new typology of linguistic contact, which fundamentally degrades the syntactic richness of the other language.

The structure of this study is as follows: Chapter 2 will follow the Introduction by addressing the two syntactic features under investigation. Chapter 3 will then detail the selection criteria for both human-authored and AI-generated texts used in the corpus construction, along with the methodology. Chapter 4 will be dedicated to the presentation and analysis of the data. Finally, Chapter 5 will conclude by drawing relevant observations.

2 Word order and pre-field occupation in German newspaper language

Although often dated, existing studies tend to converge, fortuitously, on a limited set of features that characterise the syntax of German newspapers. Unsurprisingly, the pursuit of uniformity across articles and publications, likely influenced by overarching guidelines for journalistic ethics and conduct such as the German Press Code (cf. Deutscher Presserat 2025), internal editorial policies (cf. for example SPIEGEL-Verlag 2021), automated revision tools (now powered by AI; cf.

[5] This means that a newspaper article may include diverse linguistic forms such as direct quotations, intertextual references, court reports, or reader letters (cf. De Cesare et al. 2016), each occupying a distinct position along the continuum between spoken and written language, or between conceptual orality and literacy (cf. Koch and Oesterreicher 1985).

Newman et al. 2023, 2024, 2025), and journalistic style guides (cf. Linden 2000; Kurz et al. 2013; Linden 2023), collectively help to standardise journalistic writing and distinguish it from other text types. Table (1) confirms the suitability of the selected sources for investigating word order, as all five address this topic. Furthermore, three of the five selected sources specifically link word order to the elements positioned in the pre-field, the syntactic slot preceding the finite verb.

Table 1: Occurrences of selected syntactic traits in the literature.

Source	Word order	Pre-field occupation
Mittelberg (1967)	✓	✓
Linden (2000)	✓	✓
Schmitt (2004)	✓	✓
De Cesare (2007)	✓	✗
Hofstätter (2020)	✓	✗

Linden (2000: 56) describes the typical structure of an informational sentence in journalistic writing as beginning with the subject, followed by the first part of the verb, relevant objects and adverbials, and finally the second part of the verb (in the case of compound tenses). Because this structure is so frequently employed, placing the subject – whether full NPs (cf. Example 1), pronouns (cf. Example 2), or subject clauses (cf. Example 3) – in the initial position constitutes the unmarked syntactic order of newspaper language (cf. Schmitt 2004: 261).[6]

(1) *Die Italiener* toben!
'The Italians are raging!' [B21; P.V.][7]

(2) *Sie* sind bulgarische Staatsbürger.
'They are Bulgarian citizens.' [B8]

(3) „*Was genau unter Antisemitismus zu verstehen ist und in welchen Situationen er vorliegt*, bleibt Gegenstand fortwährender wissenschaftlicher und gesellschaftlicher Reflexion; [...]."
'Exactly what is to be understood by antisemitism, and in which situations it occurs, remains the subject of ongoing scientific and societal reflection.' [SZ19]

[6] Here 'markedness' refers to the asymmetry between two or more structural variants. The unmarked form represents the default, neutral, or most expected option, which requires no specific pragmatic motivation.

[7] Each sentence in the corpus is annotated with a source label and a corresponding article number, followed by my own translations (see Abbreviations for details). The words in italics are used to emphasise the entire constituent that precedes the verb in the pre-field.

From a statistical standpoint, Wöllstein (2022: 57) reports that the subject occupies the pre-field position in approximately 50–55% of main clauses. However, while +SVO (subject-verb-object) order is expected to be predominant, it should not be assumed to be the exclusive word order.[8]

In fact, whilst the pre-field, a syntactically prominent position, crucially requires the presence of exactly one constituent, the form and complexity of this constituent can vary. As Wöllstein (2022: 55) observes, "alle Satzglieder vorfeldfähig sind" ['all sentence elements are capable of occupying the pre-field'; P.V.]. Consequently, the subject is in structural competition with other constituents such as temporal and spatial adverbials, or object phrases that can also initiate the sentence. Example (4) exhibits a causal adverbial, (5) a direct object, (6) an indirect object, and (7) a predicate.

(4) *Aus zwei Gründen* waren die Sätze aber doch bemerkenswert, […].
'For two reasons, the sentences were nevertheless remarkable.' [SZ8]

(5) *Zusätzliche Brisanz* hat das Thema erhalten, […].
'Additional explosive potential has been added to the topic.' [SZ19]

(6) *Ihm* bleibt also das Schicksal der Innenministerin […] erspart, […].
'He is thus spared the fate of the Interior Minister.' [SZ4]

(7) *Begonnen* hat das Drama am Sonntag.
'The drama began on Sunday.' [B1]

Syntactic variation within newspaper discourse is therefore both inevitable and, arguably, functionally necessary. This observation is corroborated by Schmitt (2004), who provides frequency data on pre-field occupancy in *BILD* and *Frankfurter Allgemeine Zeitung* (FAZ), as detailed in Table (2). Furthermore, findings by Wöllstein (2022: 56–57) indicate that adverbials occupy the pre-field in approximately 35% of cases, with other constituents following at around 20%.

[8] This study employs a binary classification of linearisation patterns to quantify the degree of divergence between human and AI-generated syntax. Throughout the following analysis, the label +SVO is operationally defined as any main clause where the grammatical subject occupies the pre-field. Although German is syntactically a V2 language, this linearisation mimics the rigid SVO order characteristic of English. Conversely, the label –SVO designates all main clauses where a non-subject constituent, such as an adverbial, object, or subordinate clause, occupies the pre-field.

Table 2: Pre-field occupation percentages in *BILD* and *FAZ*.

Typology	*BILD* (%)	*FAZ* (%)
direct object	18.8	14.6
indirect object	7.1	6.1
prepositional phrase	8.6	13.5
predicate noun	8.1	7.2
parts of the predicate	5.0	4.3
finite verb	2.9	0.1
particle	1.4	1.2
adverbial phrase	45.8	49.1
miscellaneous	2.3	3.9
Σ	100	100

This variability in sentence structure stems from the dynamic influence of the sentence's information structure on the selection of the constituent occupying the pre-field (cf. Wöllstein 2022: 57). For instance, while the topicalisation of the object (both direct, cf. *supra* Example 5, and indirect, cf. *supra* Example 6) or the predicate (cf. *supra* Example 7) is considered by Mittelberg (1967: 272–273) "Ausdruck gefühlsbetonter Syntax" ['a manifestation of emotionally charged syntax'; P.V.], initial temporal and spatial adverbials function to ground the event in time and space or ensure thematic continuity with the preceding discourse. Consequently, although displacing the subject from the pre-field is generally deemed a non-standard practice in journalistic writing, the occupation of the pre-field by frame-setting adverbials constitutes a standard text-grammatical strategy (cf. Weinrich 2007).

3 The corpus

The construction of the corpus is governed by the methodological principle of functional equivalence. To analyse linguistic interference effectively, it is essential to ensure that the human-authored and AI-generated texts are comparable in terms of genre (political reporting), register (formal, information-oriented standard German), and communicative intent. The overarching rationale is to minimise external variables, such as stylistic idiosyncrasies, dialectal variation, or low-quality drafting, that could obscure the specific syntactic patterns under investigation. Consequently, the selection criteria detailed below were chosen to align the *gold standard* of codified German media with the current *state-of-the-art* capabilities of LLMs, ensuring that observed divergences can be attributed to the underlying production mechanisms (human cognition vs. statistical probability) rather than disparities in task complexity or text type.

This study utilises stringent selection criteria to ensure the relevance and representativeness of its data sources, considering the inherent infinitude of LLM output. Specific criteria were systematically applied to three key components: (i)

the newspaper articles under analysis, (ii) the LLMs utilised, and (iii) the composition and size of the constructed corpus. A comprehensive explanation of these methodological parameters will be presented in the following section, providing transparency regarding the study's data selection process.

3.1 Human-authored articles

As primary criterion, circulation metrics were considered; however, these are not used here as a proxy for normative linguistic representativeness, but rather for sociolinguistic impact. By selecting the most widely circulated publications, we ensure that the corpus reflects the varieties of written German that have the highest reception among the general public. Thus, the corpus represents the *mainstream media* effectively, encompassing both the high-register norms of quality press and the compressed, emotive style of the tabloid press. Table (3) below presents distribution data for the five highest-circulation newspapers in Germany, based on figures published by the *Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V.* (IVW) for the third quarter of 2024.

Table 3: Q3 2024 newspaper circulation figures.

Newspaper	Copies
BILD	1,381,554
Süddeutsche Zeitung	261,024
Rheinische Post	206,152
Frankfurter Allgemeine Zeitung	182,763
Westdeutsche Allgemeine Zeitung	69,930

While circulation metrics offer a robust method for selecting pertinent newspapers, an additional selection criterion pertains to the publication format, distinguishing between tabloids and quality newspapers (see Table 4).

Table 4: Highest-circulation newspapers in Germany and publication format.

Newspaper	Format
BILD	Tabloid
Süddeutsche Zeitung	Quality
Rheinische Post	Quality
Frankfurter Allgemeine Zeitung	Quality
Westdeutsche Allgemeine Zeitung	Quality

If the tabloid press represents a distinct category characterised by a sensationalist style, concise texts, and emotionally charged reporting (cf. Lünenborg 2013; Raabe 2013a), quality newspapers are defined by more complex textual structures and a

reporting style that prioritises depth, accuracy, and an information-oriented approach (cf. Raabe 2013b). For this reason, one newspaper for each category has been selected.

According to Meyn (1985), a newspaper qualifies as national if more than 15% of its circulation is distributed outside its core regional area. Newspapers selected for analysis are national in audience reach, offering both national and international coverage (cf. Table 5).

Table 5: Highest-circulation newspapers in Germany and audience reach.

Newspaper	Reach
BILD	National
Süddeutsche Zeitung	National
Rheinische Post	Regional
Frankfurter Allgemeine Zeitung	National
Westdeutsche Allgemeine Zeitung	Regional

Following the establishment of criteria for newspaper selection, the subsequent phase involved the careful curation of individual articles for inclusion in the corpus. The main selection criterion pertains to the quantity of articles extracted from each newspaper. A standardised sample size of 25 articles per publication was established. While this sample size precludes broad statistical generalisation across the entire German media landscape, it provides a sufficient data density for qualitative syntactic analysis and the identification of recurrent interference patterns.

The second criterion relates to the thematic focus. For this study, articles were selected exclusively from the *Politik* 'Politics' section. This choice was motivated by the need for thematic consistency, which enhances the internal homogeneity of the corpus.

The third criterion addresses authorship. De Cesare et al. (2016: 107) distinguish four types of article authorship in online newspapers, based on their level of dependence on external sources:

(i) articoli ripresi tali e quali dalle agenzie (si tratta per lo più di articoli non firmati, che possono essere anche relativamente lunghi; [...]);
(ii) articoli con un grado molto basso di riscrittura da parte del giornalista (in questo caso, si indica semplicemente la fonte delle informazioni: nome del giornale e/o nome della o delle agenzie di riferimento);
(iii) articoli con un grado più elevato di rielaborazione, la cui confezione si avvale di fonti diverse (sono articoli firmati dalla redazione online; [...]);
(iv) articoli molto più personali e legati al nome del giornalista che firma il pezzo (in questo caso, gli articoli sono firmati con il nome del solo giornalista [...]).

(i) ‘articles reproduced exactly as received from news agencies (these are mostly unsigned articles, which can also be relatively long; […]);
(ii) articles with a very low degree of rewriting by the journalist (in this case, the source of the information is simply indicated: name of the newspaper and/or name of the relevant news agency or agencies);
(iii) articles with a higher degree of reworking, drawing on various sources (these articles are signed by the online editorial staff); [...]
(iv) more personal articles linked to the name of the journalist signing the piece (in this case, the articles are signed with the journalist’s name alone [...]).’ [P.V.]

In the selected newspapers, the classification of articles according to the typology proposed by De Cesare et al. (2016) can be substantially simplified due to differences in transparency regarding authorship and sourcing. On the one hand, *BILD* demonstrates a general lack of transparency in this regard. While many articles are signed and could, therefore, be assumed to correspond to type (iv), unsigned articles provide no clear indication of their origin. As a result, these articles could plausibly fall into any of the categories (i) to (iii), making their classification unreliable. On the other hand, *SZ* offers a more consistent and explicit approach to source attribution. It publishes both signed articles and pieces that clearly indicate their origin from established news agencies such as Deutsche Presse-Agentur (dpa), Reuters, or Bloomberg. However, even in these cases, determining the extent to which such articles have been reworked or edited remains problematic and introduces a variable of unknown agency. Consequently, only articles demonstrably authored by in-house journalists have been included. This ensures a focus on original newspaper language rather than on texts mediated through external sources.

Lastly, only informative articles have been selected for analysis. This category includes journalistic pieces whose primary purpose is to convey factual information such as data, reports, or accounts of political events and statements. Opinion pieces, editorials, and columns have been excluded, as their subjective and interpretative nature diverges from the study’s focus on political news discourse. Similarly, typical tabloid-style articles, which prioritise sensationalism over information, were also omitted, as they do not align with the objective of analysing political reporting in a more traditional informational format.

3.2 AI-generated texts

Leaderboards have become a central tool for assessing and comparing the capabilities of LLMs in a transparent and reproducible manner. For the purposes of this study, two leaderboards were used as reference points to assess the capabilities of LLMs.

One is LiveBench, a benchmark-based leaderboard that provides detailed evaluations of linguistic competence using structured prompts and reference answers (cf. Table 6). This kind of leaderboard is characterised by the use of static

and predefined datasets designed to test specific model capabilities within clearly defined categories, such as language comprehension, reasoning, instruction following, coding, and data analysis (cf. White et al. 2024). Regarding the context of language, “they include multiple language comprehension tasks”, such as “completing word puzzles, [...] fixing misspellings but leaving other stylistic changes in place, and [...] reordering scrambled plots of unknown movies”. (cf. White et al. 2024: 6) To ensure completeness and transparency, models that ranked lower in either leaderboard were also included to ensure a representative range of architectures (open vs. proprietary), as outlined subsequently.

Table 6: Ranking metrics on LiveBench.[9]

Rank	LLM	Score
1	o1-2024-12-17	66.02
2	claude-3-5-sonnet-20240620	59.80
3	o1-mini-2024-09-12	59.09
4	gpt-4o-2024-08-06	56.03
5	gemini-1.5-pro-002	54.94
6	chatgpt-4o-latest-0903	54.25
7	gpt-4o-2024-05-13	53.98
8	gemini-1.5-pro-exp-0827	53.75
9	meta-llama-3.1-405b-instruct-turbo	53.24
10	qwen2.5-72b-instruct	52.89
	[…]	
16	mistral-large-2407	48.50

The other is LMArena (formerly LMSYS and Chatbot Arena), which aggregates human preference data to rank models based on real-world user evaluations (cf. Table 7). This allows users to interact with the models in real-time and provide informed judgments regarding the quality of the generated responses, with over 100 languages represented in the gathered data (cf. Zheng et al. 2024).

[9] For clarity, the numerical values accompanying model names often denote the release date (e.g., o1-*2024-12-17*) or the parameter count (e.g., qwen2.5-*72b*-instruct, signifying 72 billion parameters). *Instruct* indicates a model specifically trained as a chatbot, while terms such as *sonnet*, *mini*, and *turbo* serve as specific monikers for the model.

Table 7: Ranking metrics on LMArena.

Rank	LLM	Score
1	ChatGPT-4o-latest (2024-09-03)	1338
2	o1-2024-12-17	1335
3	o1-mini	1314
4	Gemini-1.5-Pro-002	1304
5	Gemini-1.5-Pro-Exp-0827	1299
6	Grok-2-08-13	1293
7	GPT-4o-2024-05-13	1285
8	GPT-4o-mini-2024-07-18	1273
9	Gemini-1.5-Flash-Exp-0827	1269
10	Claude 3.5 Sonnet	1268
	[…]	
11	Meta-Llama-3.1-405b-Instruct-bf16	1267
18	Mistral-Large-2407	1251

Another significant distinction concerns the difference between (i) proprietary and (ii) open-source models. Proprietary LLMs are developed, maintained, and operated exclusively by private companies. They are typically accessible via commercial platforms, often as paid services or within the constraints of a limited free tier. Although users retain control over prompt design, many proprietary systems do not disclose critical information such as system prompts, hyperparameters, or architectural details, thereby limiting external scrutiny and reproducibility. In contrast, open-source language models are publicly released and freely available for download. Users can run them locally or in self-managed cloud environments, gaining full control over the model's deployment and configuration. This study incorporates two proprietary and two open-source language models, with the constraint of selecting only one model per company.

Lastly, the selection was guided by two further criteria: (i) the inclusion of the most recent publicly available version of each model, and (ii) the preference for models with the highest number of parameters, where such information was available. The rationale for selecting the latest version is grounded in the assumption that newer iterations incorporate improvements in training data quality, model architecture, and optimisation techniques. Similarly, the number of parameters serves as a proxy for architectural complexity, which is generally associated with a model's capacity to learn nuanced representations and generalise across diverse tasks. However, this metric is not always disclosed, particularly in the case of proprietary models. Table (8) summarises these criteria.

Table 8: LLM selection criteria.[10]

Model	Rank	License	Company	Successor[a]	Param. (B)
o1	2.5	Closed	OpenAI	–	–
Claude 3.5 Sonnet	6	Closed	Anthropic	–	175
llama-3.1-405b-instruct	10	Open	Meta AI	–	405
mistral-large-2407	17	Open	Mistral AI	–	123

When using LLMs to produce precise and contextually appropriate responses, prompting strategies play a crucial role. For the purposes of this study, a *zero-shot* prompt with *chain-of-thought* reasoning was employed, which is a format widely used in benchmark evaluations due to its ability to reveal the model's inherent capabilities without relying on example-based learning.[11] However, the success of this strategy is closely tied to the clarity and precision of the linguistic instructions given. To this end, the prompt was constructed following the principles of prompt engineering and based on a tripartite structure consisting of (i) a clearly defined role, (ii) a specific task, and (iii) an expected output format (cf. Phoenix and Taylor 2024). In addition, information was explicitly categorised and visually delimited to improve the model's comprehension of prompt structure. According to OpenAI (2025), using delimiters such as ###, triple quotes (""""), or tags (<tag>) can help separate and clarify distinct components of the prompt, thereby reducing potential confusion and increasing response coherence. The following prompt shows the original German text alongside the English translation.[12]

```
ROLLE: Du bist eine Person mit professioneller Erfahrung
im Bereich der politischen Analyse. Täglich verfasst du
eine Kolumne, die sich auf politische Themen und
Ereignisse konzentriert.
###
AUFGABE: Verfasse einen Artikel über ein aktuelles
politisches Ereignis. Gehe dabei Schritt für Schritt
vor.[13]
###
ZIEL: Erstelle einen Text, der in einer journalistischen
Kolumne veröffentlicht werden kann.
```

[10] At the time of analysis.

[11] *Zero-shot* generation occurs when the LLM creates output without being provided any examples. In contrast, *chain-of-thought* is a strategy, initially activated by specific phrases like "Proceed step by step", but now often integrated into the models, that prompts the LLM to process information internally before delivering the final requested output.

[12] I would like to thank Professor Anna-Maria De Cesare Greenwald and Michela Gargiulo for their observations regarding gender in the original prompt design. The prompt now utilises an impersonal point of view to *prevent* gender bias in the output.

[13] In the case of model o1, the chain-of-thought was omitted, as the model is said to have this reasoning mechanism integrated natively, rendering the instruction redundant.

> ‘ROLE: You are a person with professional experience in the field of political analysis. You write a daily column focusing on political topics and events.
> ###
> TASK: Write an article about a current political event. Proceed step by step.
> ###
> GOAL: Create a text suitable for publication in a journalistic column.’ [P.V.]

The prompts were kept identical across all models. No model-specific prompt engineering or system-prompt adjustments were made to accommodate specific architectures or token limits. This *zero-shot, zero-modification* approach serves to evaluate the models’ baseline performance in handling the requested task, treating the underlying training data as the sole independent variable.

All texts were generated via the standard public web interfaces (WebUIs) of the respective models in December 2024. Consequently, the generation parameters (Temperature, Top-P, Top-K) correspond to the default system configurations set by the developers at the time of access. While API-based generation allows for granular control (e.g., forcing Temperature to 0), the use of the WebUI was a deliberate methodological choice to ensure ecological validity. This approach assesses the models’ linguistic competence as it is exposed to and utilised by the vast majority of human users, rather than in a theoretical laboratory setting. Moreover, to control for stochastic variability and history bias, each text was generated in a discrete, isolated chat session. The context window was reset (i.e., *New Chat*) for every single generation. This ensures that no residual information or stylistic priming from previous outputs influenced subsequent generations, guaranteeing a *tabula rasa* state for each prompt.

25 texts per LLMs were generated. The size was determined based on the principle of qualitative saturation rather than statistical generalisation (cf. McEnery and Hardie 2012). Preliminary analysis indicated that structural interference patterns in LLM output tend to be highly repetitive; thus, this sample size provides sufficient density to categorise these interference types without redundancy. Table (9) presents a quantitative summary of the assembled corpus.

Table 9: Corpus composition.

Category	Source	Sentence Count
Human-authored	*BILD*	665
	Süddeutsche Zeitung	1,117
AI-generated	Total Human	1,782
	o1	383
	Claude 3.5 Sonnet	537
	llama-3.1-405b-instruct	356
	mistral-large-2407	510
	Total AI	1,786
	Total	3,568

3.3 Methodology

Quantitative analysis was conducted using the Pearson chi-square test of independence to evaluate distributional differences between human and AI-generated datasets. All statistical computations were performed in the R computing environment (v.4.5.2) via RStudio (2025.09.2) using the standard *chisq.test* function. Given the adequate sample size across all contingency tables, Yates' continuity correction was disabled (*correct = FALSE*).

A strict significance level of $\alpha = 0.01$ was adopted ($p < 0.01$), rather than the standard $\alpha = 0.05$. This threshold was selected to ensure robustness against the large token count of the corpus, where minor fluctuations might otherwise yield statistical significance. This approach mitigates the risk of Type I errors (False Positives) inherent in multiple hypothesis testing without requiring overly conservative corrections that might obscure genuine mid-sized effects.

Since the chi-square test assesses difference rather than similarity, a supplementary equivalence criterion was defined. We established a region of practical equivalence with a tolerance margin of ±10 percentage points (pp) relative to the human baseline. Formally, for a given feature frequency P, equivalence is accepted if: $P_{Human} - 0.10 \leq P_{AI} \leq P_{Human} + 0.10$. This margin acknowledges that linguistic style is inherently variable; a stricter bound (e.g., ±5%) would be overly sensitive to natural authorial variation, while a wider bound would fail to capture significant structural interference. Thus, *equivalence* in this study denotes adherence to the target genre's frequency norms within a pragmatic tolerance window.

Following the quantitative analyses of human-authored newspaper articles and AI-generated texts, a detailed, albeit largely manual, linguistic annotation process was conducted.

4 Results

Illustrative examples from the human-authored articles corpus demonstrate the expected classification categories: +SVO (cf. Example 8), –SVO (cf. Example 9), and non-sentential units (cf. Example 10).[14]

(8) *Das britische Heer* hat 500 Boxer bestellt.
'The British army has ordered 500 Boxer vehicles.' [SZ24]

(9) *Dagegen* hatten einige Länder Widerstand angekündigt.
'By contrast, several countries had announced their opposition.' [B3]

(10) Phase 3 aus der „Verbreitung des Narrativs über soziale Netzwerke".
'Phase 3: "Dissemination of the narrative via social networks".' [B7]

The subsequent comparative figure illustrates the distribution of these structures in the corpus of human-authored articles (cf. Figure 1).[15] The present data seem to challenge the assumption that +SVO constitutes the predominant structural pattern in German newspaper syntax. Instead, these findings align with prior observations suggesting that both SVO and non-SVO constructions are employed strategically to enhance textual variety and reader engagement.

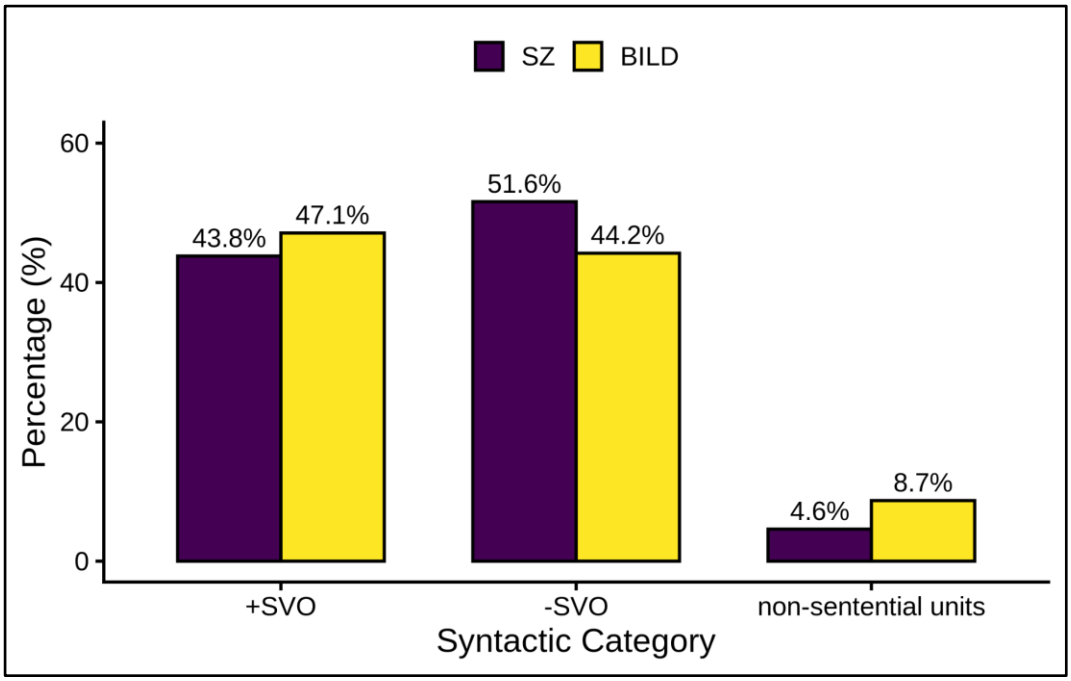

Figure 1: Comparison of word order in *SZ* and *BILD*.

[14] Following Wöllstein (2022)'s terminology, utterances lacking full sentential structure, typically due to predicate omission, are classified as *Nicht-Sätze* ('non-sentences').
[15] Absolute figures are provided in Table A of the Appendix.

By comparison, output generated by the LLMs exhibits notable inconsistencies. Despite receiving identical prompt instructions, the models produce varied sentence structure patterns (see Figure 2).

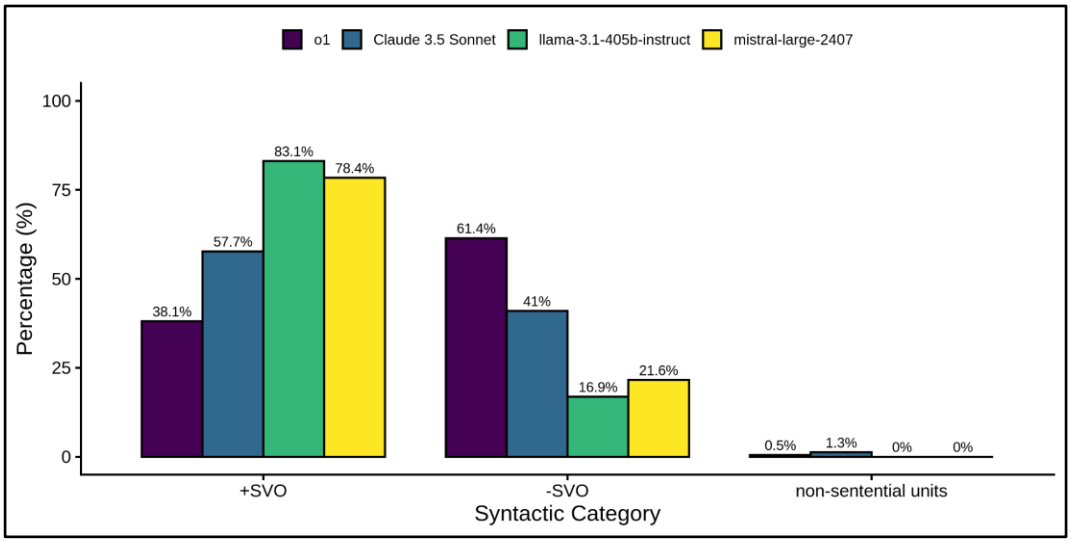

Figure 2: Comparison of word order between LLMs.

Examples illustrating the different structural types follow: a +SVO sentence in (11), a –SVO construction in (12), and a non-sentential unit in (13).

(11) *Dieser Plan, bekannt als der „Green Deal“,* umfasst eine Vielzahl von Maßnahmen, […].
‘This plan, known as the “Green Deal,” includes a wide range of measures.’ [M16]

(12) *Andererseits* bedeutet der Brexit auch eine Herausforderung für die EU.
‘On the other hand, Brexit also presents a challenge for the EU.’ [L3]

(13) Eine Ära, in der alt Gewissheiten hinterfragt werden und neue Antworten gefunden werden müssen.
‘An era in which old certainties are being questioned, and new answers must be found.’ [C19]

Notably, both llama-3.1-405b-instruct and mistral-large-2407 generated no non-sentential units in their outputs. In contrast, Claude 3.5 Sonnet and o1 produced a limited number of non-sentences, though these quantities remain substantially lower than those observed in the human-authored corpus.

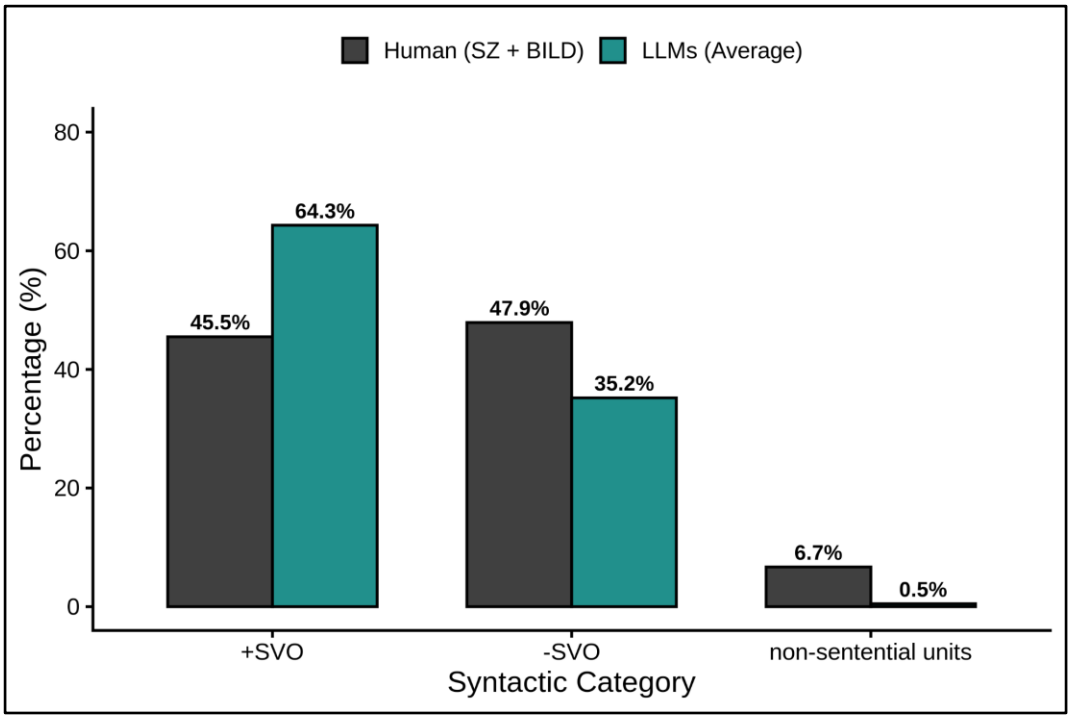

Figure 3: Comparative distribution of syntactic linearisation patterns.

Figure (3) synthesises the distributional data from the human and AI-generated sub-corpora, allowing for a direct comparison of linearisation strategies. While the human baseline exhibits a roughly balanced distribution between subject-initial (45.5%) and non-subject-initial (47.9%) constructions, the LLMs display a pronounced skew toward subject rigidity. Specifically, the AI-generated texts show a +SVO frequency of 64.3%, exceeding the human benchmark by approximately 19 percentage points. Conversely, the usage of the pre-field for non-subject constituents drops to 35.2% in the AI output. This indicates that while the models can generate proper German V2 structures, they significantly underutilise the inherent topological flexibility of the language.[16] Instead, they converge on a repetitive SVO pattern that aligns more closely with English syntactic norms than with the pragmatic variation typical of German newspaper language.

To assess these findings quantitatively, chi-square tests were performed comparing the sentence structure distributions in each model-generated corpus with the aggregated human-authored corpus (Σ_H). The statistical results are detailed in Table (10).

[16] A verb-second (V2) language, like German, is characterised by the finite verb's obligatory position as the second constituent, a requirement that applies either specifically to main clauses or across all finite clauses (Holmberg 2015).

Table 10: Comparison of LLMs and Human baseline using chi-square test results.[17]

Model	χ^2	*p*-value	Result
o1	12.08	< 0.001	Significant
Claude 3.5 Sonnet	16.60	< 0.0001	Significant
mistral-large-2407	141.02	< 0.0001	Significant
llama-3.1-405b-instruct	141.87	< 0.0001	Significant

The extremely low *p*-values confirm statistically significant differences in the distribution of +SVO and –SVO structures between human-authored and AI-generated texts. However, the data also reveals distinct patterns among the models. Whilst llama-3.1-405b-instruct and mistral-large-2407 predominantly generate +SVO sentences (83.1% and 78.4%, respectively), and Claude 3.5 Sonnet maintains a more balanced distribution between +SVO and –SVO structures, o1 produces a higher proportion of –SVO sentences than found in human-authored articles.

Formal equivalence testing is needed to determine if these model outputs signify meaningful differences or merely exceed acceptable variation. This testing would require maintaining a ±10% tolerance, which defines an equivalence range of [38.2%; 58.2%] based on aggregated human texts (see Table 11).[18]

Table 11: Equivalence in word order.

Source	+SVO (%)	Δ	EQ	≈
Σ_H	48.2	0,00	Ref.	–
o1	38.3	-9.9	38.3 ∈ [38.2; 58.2]	✓
Claude 3.5 Sonnet	58.5	+10.3	58.5 > 58.2	✗
llama-3.1-405b-instruct	83.1	+34.9	83.1 >> 58.2	✗
mistral-large-2407	78.4	+30.2	78.4 >> 58.2	✗

Only model o1 falls within the defined equivalence range, indicating statistical equivalence to human texts regarding the proportion of +SVO sentences. It is important to note, however, that this value is precisely at the boundary of the acceptance range, a position similar to that occupied by Claude 3.5 Sonnet.

Focus now shifts to the formal types of subjects that appear in the initial sentence position. The corpus of human-authored articles reveals the following

[17] For all tests, the degree of freedom (*df*) was 1. Units that were not complete sentences were excluded from the analysis.

[18] As this is a binary property, the analysis can focus on either +SVO or –SVO occurrences. In this case, +SVO was examined.

subject types: full NPs (cf. Example 14), pronouns (cf. Example 15), and subject clauses (cf. Example 16).

(14) *Diese Zahl* nennt die staatliche Ombudsstelle für Bürgerrechte auf ihrer Homepage.
‘The state ombudsman’s office for civil rights cites this number on its homepage.’ [SZ16]

(15) *Sie* werde eine Präsidentin „für alle Amerikaner“ sein.
‘She said she would be a president “for all Americans”.’ [B5]

(16) *Ob nun der Kanzler angesichts solcher Wachstumszahlen mit frischen Ideen zurückkehrt, um das Haushaltspuzzle zu vervollständigen,* ist die große Frage.
‘Whether the Chancellor will now return with fresh ideas to complete the budget puzzle in light of such growth figures is the big question.’ [SZ1]

Upon examining Figure (4), it becomes clear that *BILD* shows a slight preference for using or reintroducing the subject as a noun rather than a pronoun. In other words, *SZ* tends to use more pronouns and subject clauses, whereas *BILD* more frequently utilises nouns and employs fewer pronouns or subject clauses.

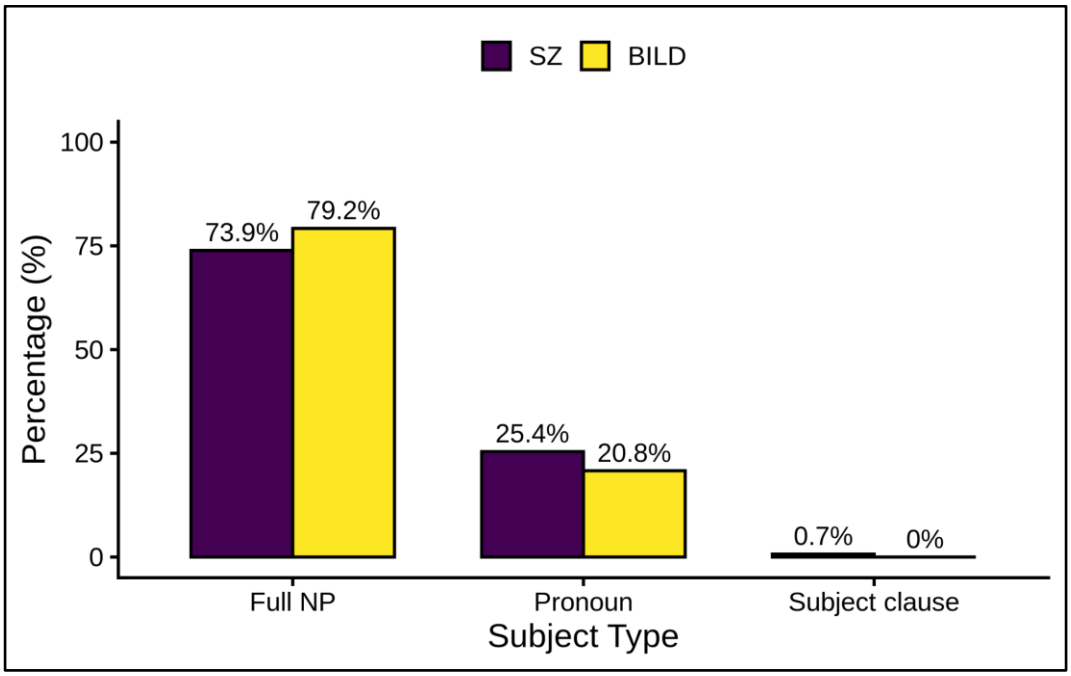

Figure 4: Comparison of subject type distribution in *SZ* and *BILD*.

However, among the LLMs, an even stronger divergence can be observed, even though the same formal subject types as in the human texts are present: a full NP in Example (17), a pronoun in (18) and a subject clause in (19).

(17) *Die EU* muss jetzt ihre Politik und ihre Institutionen an die neue Realität anpassen und sicherstellen, dass sie weiterhin effektiv und relevant bleibt.
'The EU must now adapt its policies and its institutions to the new reality and ensure that it remains effective and relevant.' [L3]

(18) *Es* ist ein schmaler Grat, auf dem sich die Politik hier bewegt.
'Politics is walking a fine line here.' [C22]

(19) *Dass sich Industrie, Wissenschaft und Politik nun zusammenraufen,* dürfte eine Blaupause für künftige Projekte sein, bei denen europäische Lösungen stärker in den Vordergrund rücken.
'The fact that industry, science, and politics are now pulling together should serve as a blueprint for future projects in which European solutions are brought more to the forefront.' [O3]

The distribution of subject types among the LLMs is presented in Figure (5).

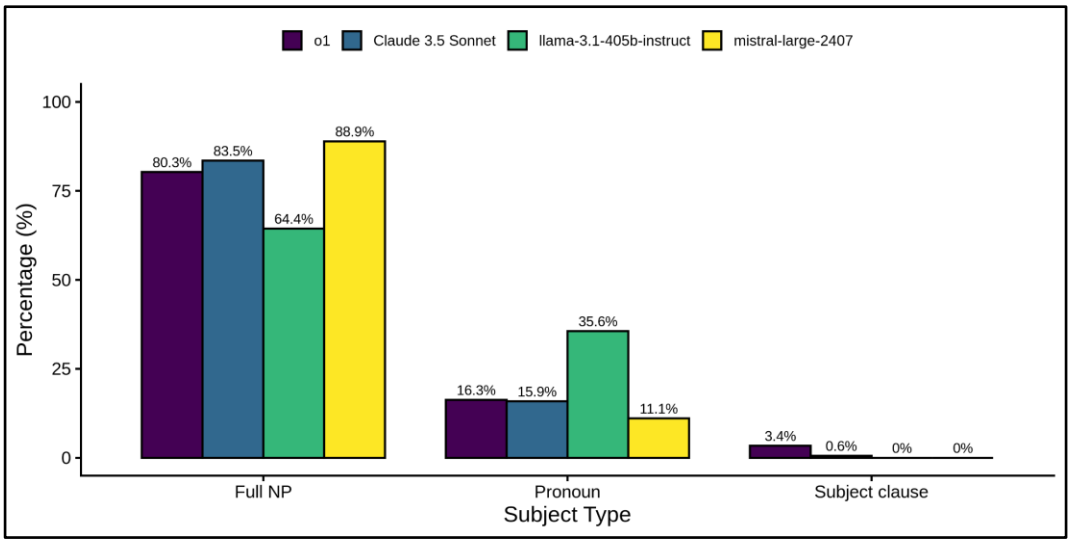

Figure 5: Comparison of subject type distribution between LLMs.

As Figure (5) clearly illustrates, certain models diverge from human texts in their formal subject distribution. For instance, mistral-large-2407 demonstrates a distinct inclination towards full NPs over pronouns in the initial position, whereas llama-3.1-405b-instruct presents the inverse pattern: in this case, pronouns occupy the initial position in 35.6% of instances.

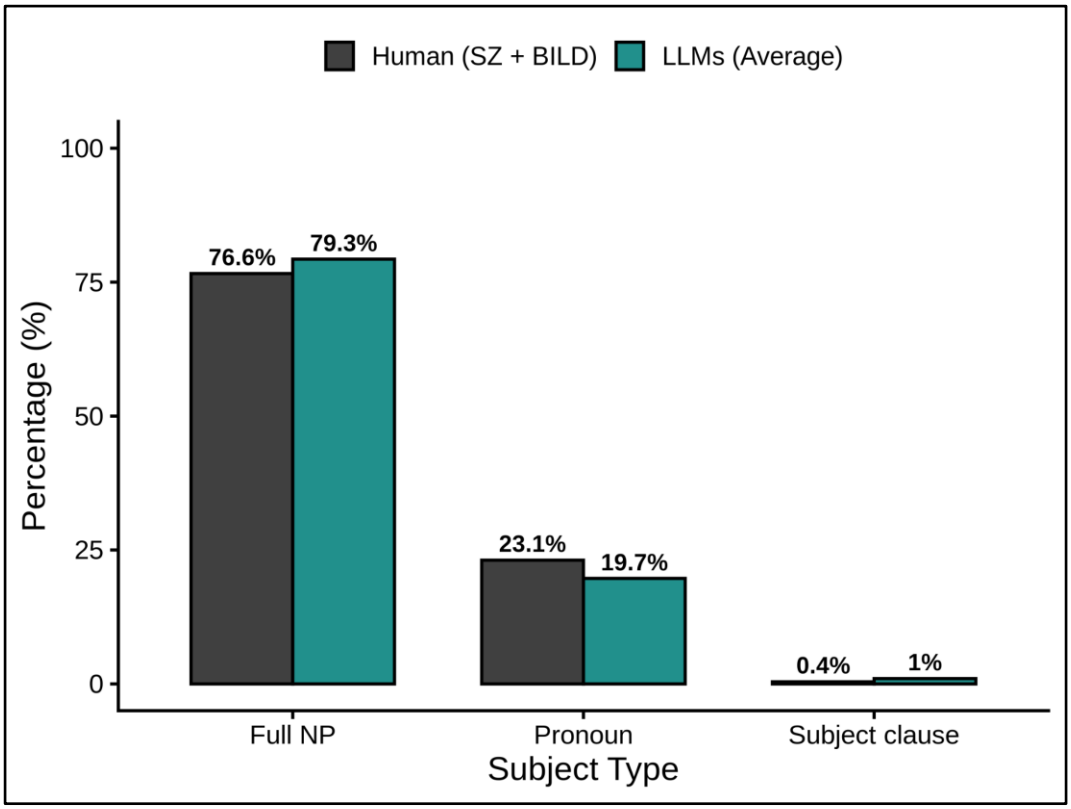

Figure 6: Comparative distribution of subject types.

However, according to Figure (6), it is observed that both human-authored articles and AI-generated texts exhibit a general preference for the use of full NPs over pronouns. This linguistic tendency can be attributed to established conventions within newspaper language, where subjects are typically maintained in their nominal form or undergo only slight variations, rather than being replaced with pronouns (cf. Linden 2000; Kurz et al. 2013; Linden 2023). This practice is primarily employed to enhance clarity and ensure the traceability of references within the text (cf. Example 20).

(20) Da wäre zuerst die Brombeerkoalition, die ein Bündnis aus CDU, SPD und dem Bündnis *Sahra Wagenknecht* (BSW) beschreiben soll. […] Die Verhandlungen sind deshalb so schwierig, weil *Wagenknecht* im Grunde Bedingungen vorgibt, [...]. Während die Thüringer BSW-Verhandler, vor allem in Person von Katja Wolf, deutlich gemacht haben, dass sie regieren wollen, bremst *Wagenknecht* aus dem Hintergrund. […] Dafür ist die Partei zu jung und zu stark auf *Wagenknecht* ausgerichtet. […] Auch für *Wagenknecht* gibt es Risiken. Löst *sie* ihr großflächig plakatiertes Wahlversprechen, für Frieden einzutreten, nicht ein, riskiert *sie* die Glaubwürdigkeit ihrer jungen Partei. [...] *Wagenknecht* hat immer wieder betont, dass das BSW auch dafür hafte, was auf Landesebene geschehe. Am

komfortabelsten wäre für *sie* die Tolerierung eines Bündnisses aus CDU und SPD in Thüringen.
'First, there would be the "blackberry coalition," a term intended to describe an alliance of the CDU, SPD, and the Sahra Wagenknecht Alliance (BSW). [...] The negotiations are so difficult primarily because Wagenknecht is essentially dictating the terms, [...]. While the BSW negotiators in Thuringia, particularly Katja Wolf, have made it clear that they want to govern, Wagenknecht is applying the brakes from the background. [...] The party is too young and too heavily focused on Wagenknecht for that. [...] There are risks for Wagenknecht as well. If she fails to deliver on her prominently featured campaign promise to champion peace, she risks the credibility of her young party. [...] Wagenknecht has repeatedly emphasized that the BSW would also be held accountable for what happens at the state level. The most comfortable option for her would be to tolerate a coalition between the CDU and SPD in Thuringia.' [SZ5]

Example (20) highlights the tendency to refer to a subject or object by repeating the same element or through slight modifications. This does not signify a complete absence of pronoun usage; however, as demonstrated in Figure (6), their application remains notably restricted. Similar observations can also be made regarding LLMs' outputs (cf. Example 21).

(21) Mit dem Sieg der pro-europäischen Opposition unter Führung von *Donald Tusk* zeichnet sich eine neue Ära ab - nicht nur für Polen, sondern möglicherweise für ganz Europa. *Tusks Triumph* ist mehr als nur ein Machtwechsel. *Er* symbolisiert die Sehnsucht vieler Polen nach einer Rückkehr zu liberalen, demokratischen Werten und einer engeren Anbindung an die Europäische Union. […] *Tusk und seine Koalitionspartner* stehen vor der Herausforderung, nicht nur zu regieren, sondern auch zu versöhnen und zu reformieren. […]
'With the victory of the pro-European opposition led by Donald Tusk, a new era is dawning - not just for Poland, but possibly for all of Europe. Tusk's triumph is more than just a change of power. It symbolizes the longing of many Poles for a return to liberal, democratic values and closer ties with the European Union. [...] Tusk and his coalition partners face the challenge not only of governing, but also of reconciling and reforming.' [C5]

While the presented data establishes the prevalence of –SVO structures in human-authored texts, quantitative and qualitative analyses reveal the functional diversity of this linearisation. Unlike the rigid subject-initial patterns observed in AI outputs, the human corpus demonstrates a flexible exploitation of the pre-field to manage information flow. The following Examples (22–26) illustrate a selection of the wide spectrum of word classes and functional constituents ranging from adverbials to object arguments that occupy the initial position in standard German

news reporting. Examples (22) through (26) illustrate several distinct patterns: a modal adverbial, a clause in the pre-field, a conjunction, a *wh*-element, and a predicative, in that order.[19] These patterns are further detailed in Figure (7).

(22) *Tatsächlich* steht der öffentliche Sektor als Gehalts-Orientierungsmaßstab im Beschlusspapier.
'In fact, the public sector is included in the policy paper as a benchmark for salaries.' [B3]

(23) *Dass Lindner damit nicht falsch liegt,* zeigt die Reaktion von SPD-Fraktionschef Rolf Mützenich auf die Steuerschätzung.
'SPD parliamentary group leader Rolf Mützenich's reaction to the tax forecast shows that Lindner is not wrong about this.' [SZ14]

(24) *Und* es scheint eine lange Zielgerade für den 81-Jährigen zu werden.
'And it seems to be a long final stretch for the 81-year-old.' [B19]

(25) *Warum* schweigt er so lange?
'Why is he silent for so long?' [SZ5]

(26) *Deutlich niedriger* ist die Zahl der Asylgesuche.
'The number of asylum applications is significantly lower.' [SZ16]

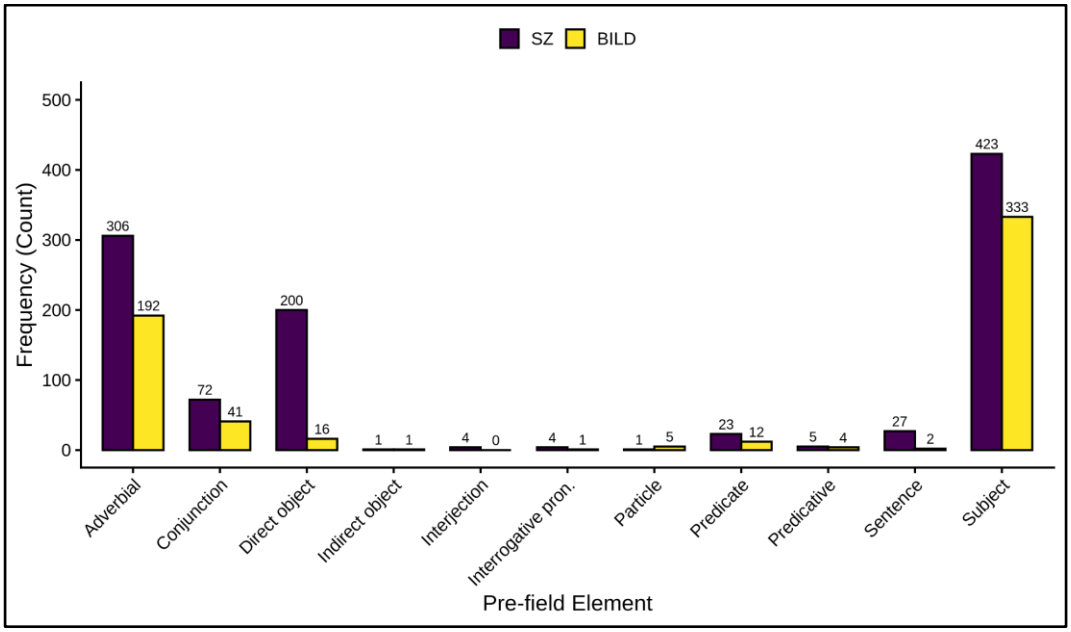

Figure 7: Comparison of pre-field occupations in *SZ* and *BILD.*

[19] Strictly speaking, conjunctions do not occupy the pre-field but rather a dedicated position; nevertheless, they constitute the first element of the sentence.

The empirical data strongly corroborates the theoretical stance that non-subject pre-field occupation constitutes a pragmatically unmarked strategy in German news discourse (cf. Schmitt 2004; Wöllstein 2022). This observation is validated by the predominance of frame-setting adverbials, which serve to anchor the narrative in time and space. Indeed, adverbial phrases account for a substantial portion of –SVO sequences: 306 of the 643 instances in SZ (47.6%) and 192 of the 274 in BILD (70.1%). The following Examples (27–28) demonstrate the semantic versatility of this topological position, featuring temporal (27), and spatial (28) adverbials.

(27) *In fast vier Wochen* ist Heiligabend.
'Christmas Eve is in almost four weeks.' [B9]

(28) *In Baden-Württemberg* konnten die Grünen ihre größten Wahlerfolge erzielen.
'The Green Party achieved its greatest electoral successes in Baden-Württemberg.' [SZ13]

Figure (8) shows distribution by adverbial type.

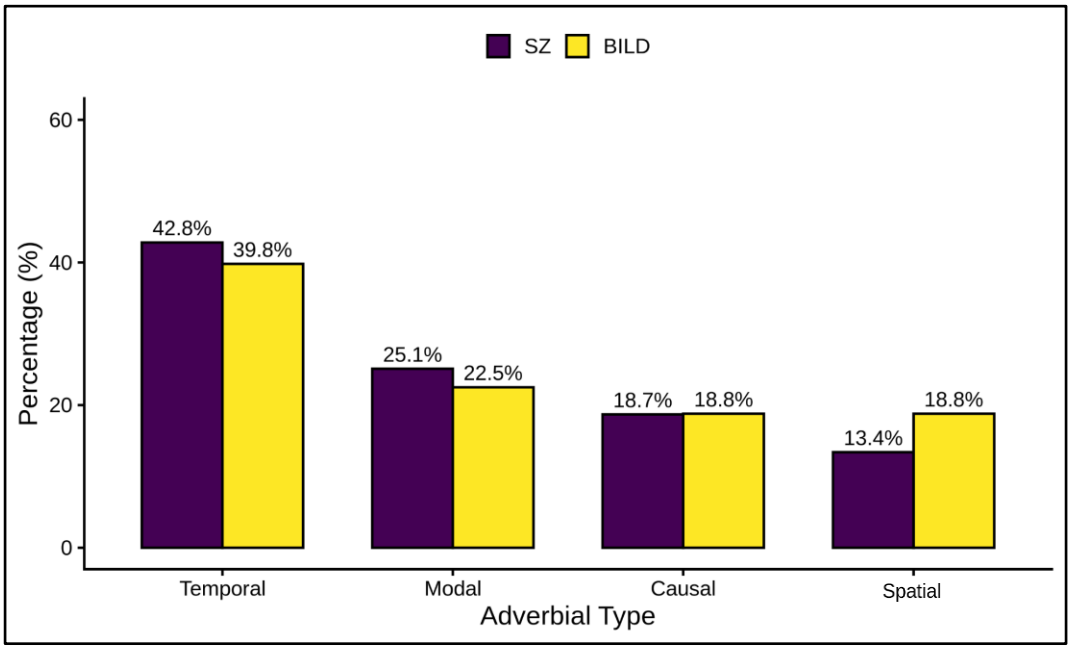

Figure 8: Comparison of adverbial by type in *SZ* and *BILD*.

Indeed, human-authored articles display a clear preference for temporal adverbials in sentence-initial position, presumably because a sentence can be described as a group of constituents that express a *state of affairs* centered around a predicate (cf. Wöllstein 2022: 27). In fact, a key distinction between a mere group of constituents and a full sentence lies in the latter's ability to temporally situate an event along a

timeline (ibid.). Temporal reference is conveyed not only through the inflection of the finite verb, but also via specific adjuncts that further specify the point in time.

The diversity of constituents found in the pre-field position is a key characteristic of human-authored articles, going beyond just the frequency of occurrence. This level of variety, however, is not entirely mirrored in AI-generated texts. For instance, Examples (29–35) illustrate a range of constituents that appear in the pre-field position within the analysed LLM corpus. Example (29) presents a modal adverbial, (30) a clause in the pre-field position, (31) a conjunction, (32) a direct object, (33) an indirect object, (34) a *wh*-element, and (35) a predicative.

(29) *Natürlich* wird von der Leyen auch vor großen Herausforderungen stehen.
‘Of course, von der Leyen will also face great challenges.’ [M14]

(30) *Ob diese Gratwanderung gelingt,* wird auch davon abhängen, […].
‘Whether this balancing act will succeed will also depend on […].’ [O13]

(31) *Aber* es erfordert auch ein hohes Maß an Verantwortung […].
‘But it also requires a high degree of responsibility […].’ [C14]

(32) *Diesen Spagat* wird die EU-Kommission weiter moderieren müssen, […].
‘The EU Commission will have to continue moderating this balancing act.’ [O7]

(33) *Ihnen* bietet die allgemeine Unzufriedenheit eine willkommene Gelegenheit, […].
‘To them, the general dissatisfaction offers a welcome opportunity.’ [O24]

(34) *Was* sind die zentralen Themen […] in diesem spannenden Wahlkampf?
‘What are the central issues […] in this exciting election campaign?’ [M25]

(35) *Besonders besorgniserregend* ist die Tatsache, […].
‘Particularly worrying is the fact […].’ (M8)

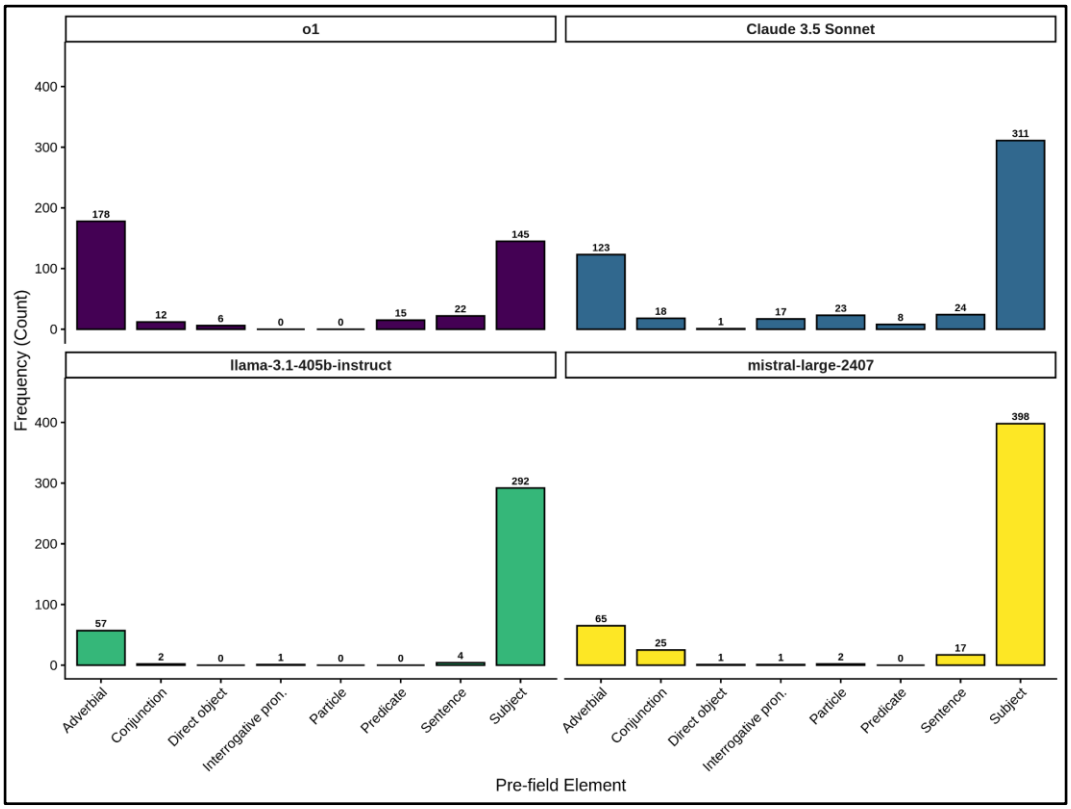

Figure 9: Comparison of pre-field occupations between LLMs.

According to Figure (9), while LLMs have globally produced a substantial variety of constituent types in the pre-field position, this trend does not appear to be universally prevalent across all models, nor does it occur with the frequency anticipated from the results of the human-authored articles corpus.[20]

LLMs, conversely, produce varied patterns in adverbial distribution. Examples (36), (37), and (38) demonstrate this by showing three different types of adverbials: temporal, causal, and spatial.

(36) *Gestern* hat die Regierung ein ambitioniertes Ziel vorgestellt, die Treibhausgasemissionen bis 2030 um 50% zu reduzieren.
‘Yesterday, the government presented an ambitious goal to reduce greenhouse gas emissions by 50% by 2030.’ [L8]

(37) *Schließlich* laufen zahlreiche Förderprojekte, die von künftigen Budgets abhängig sind.
‘After all, numerous funding projects are underway that depend on future budgets.’ [O24]

[20] Three categories with negligible counts across all LLMs (i.e., interjections, indirect objects and predicatives) were omitted from this figure to improve readability.

(38) *An der Tankstelle* dürfte der Liter Benzin um etwa 7 Cent teurer werden, Diesel um etwa 8 Cent.
'At the pump, a litre of petrol is expected to become about 7 cents more expensive, and diesel about 8 cents.' [C22]

Figure (10) provides a comprehensive visual representation of the distribution patterns observed across different adverbial types within the four LLMs under investigation.

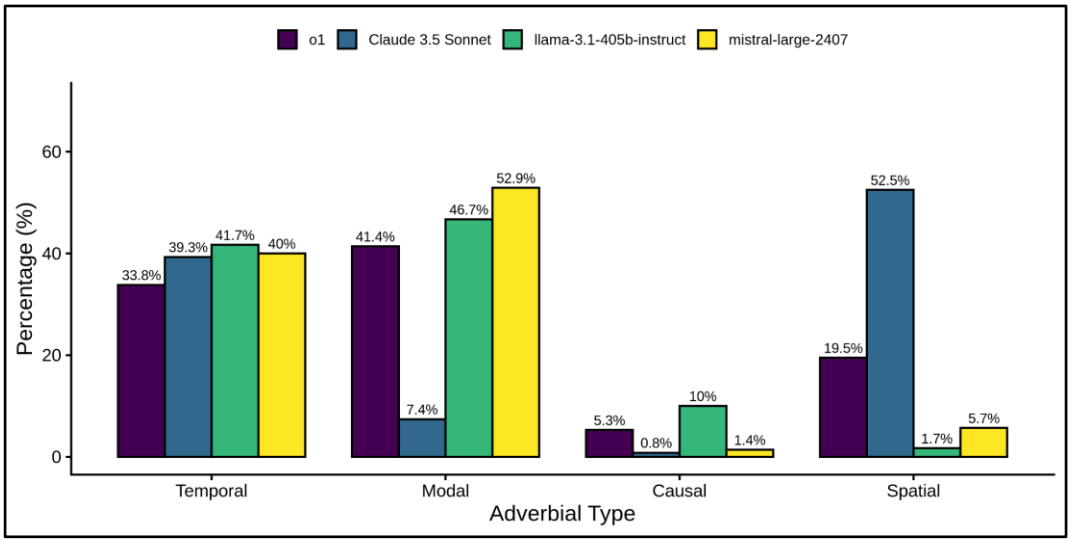

Figure 10: Comparison of adverbial by type between LLMs.

While LLMs demonstrate a substantial presence of sentence-initial temporal adverbials (146 instances), they exhibit an overall preference for modal adverbials, which constitute the most frequent type in initial position across all models (184 instances). In contrast, both spatial and causal adverbials appear significantly less frequently in AI-generated texts compared to the human-authored corpus.

Figure (11) details the specific grammatical categories occupying the pre-field in both sub-corpora.[21] The comparative analysis highlights a significant typological gap: while human authors frequently utilise the pre-field for direct objects (12.9%) and adverbials (29.7%) to manage information focus and thematic progression, LLMs show a marked deficiency in object-fronting (dropping to a negligible 0.5%). Instead, when the models do deviate from the subject-initial norm, they restrict their non-subject initial sentences almost exclusively to adverbials (24.4%), failing to replicate the full spectrum of German topological options.

[21] To enhance readability, the three categories with negligible counts across newspapers and LLMs (specifically, interjections, indirect objects, and predicatives) were excluded from this figure.

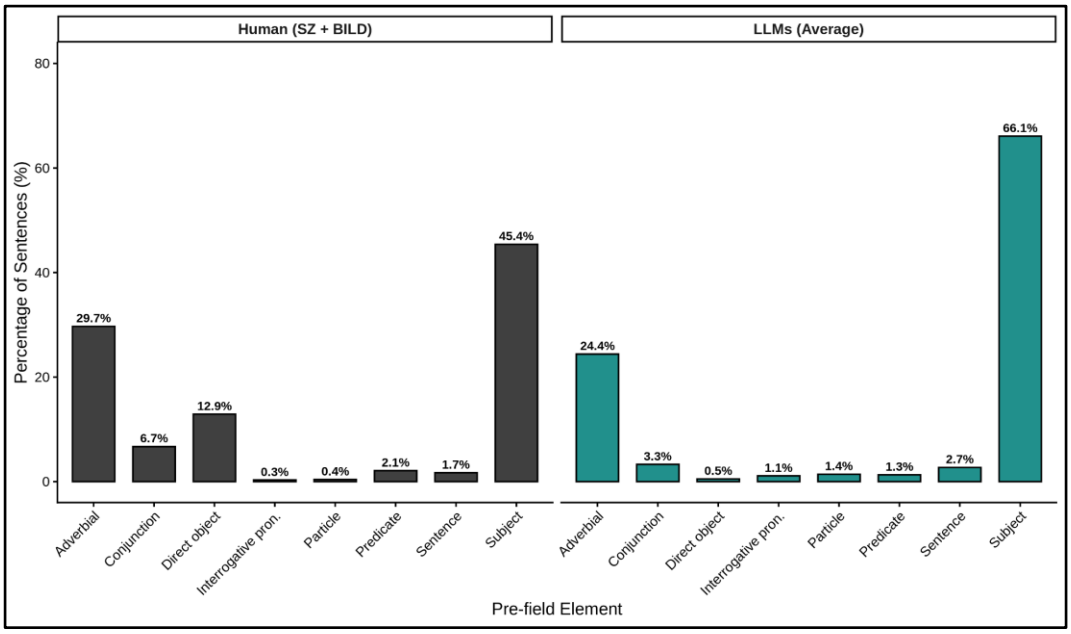

Figure 11: Comparative distribution of pre-field elements.

Given the variability in part-of-speech usage for pre-field positioning across different models, an equivalence test can also be applied to this aspect. This analysis enables us to ascertain whether the models demonstrate the same creative diversity in occupying the initial sentence position as human-authored texts (cf. Table 12).

Table 12: Equivalence in pre-field occupation.

Source	PF (%)	Δ	EQ	≈
Σ_H	100.0	0.00	Ref.	–
o1	72.7	-27.3	72.7<90.0	✗
Claude 3.5 Sonnet	81.8	-18.2	81.8<90.0	✗
llama-3.1-405b-instruct	45.5	-54.5	45.5<<90.0	✗
mistral-large-2407	72.7	-27.3	72.7<90.0	✗

The quantitative analysis of the models' performance against human reference reveals a consistent pattern: none of the examined LLMs achieved the predefined equivalence threshold. This finding underscores a considerable disparity between AI-generated text and human-authored content concerning the evaluated linguistic features. o1 and mistral-large-2407 exhibited identical performance, achieving 72.7% of the human reference. This translates to a substantial 27.3-percentage point deviation below the minimum equivalence threshold. The identical performance suggests that these models may share similar underlying architectural characteristics and training data biases, resulting in comparable limitations, despite alleged differences in the linguistic distribution between English and German training data. The model that demonstrated the closest approximation to human-

authored text, achieving 81.8%, was Claude 3.5 Sonnet. Although this represents the highest score among the models tested, it nonetheless constitutes an 18.2-percentage point deficit from the equivalence standard established by the corpus of human-authored articles. Lastly, llama-3.1-405b-instruct displayed the most significant discrepancy, reaching only 45.5% of the human reference, which represents a considerable 54.5-percentage point deviation, indicating a pronounced difficulty in replicating the linguistic nuance observed in human-authored texts.

Across all models, a clear and consistent trend emerges. The examined LLMs generally exhibit reduced diversity in pre-field constituent placement when compared to authentic German journalistic texts. The pre-field, a prominent position in German sentence structure, offers considerable flexibility for stylistic variation and emphasis in human writing. The diminished creative variation observed in the LLM outputs suggests that these models tend towards more conventional or simplified syntactic arrangements, failing to fully exploit the stylistic possibilities inherent in German grammar. While the degree of this deviation varies notably among the individual models, the overarching pattern of diminished creative variation in pre-field positioning relative to human-authored texts remains constant. This consistent shortcoming has significant implications for the naturalness and stylistic richness of AI-generated German text, particularly in contexts where nuanced expression and varied sentence structures are crucial, such as journalism, literature, or academic writing.

In summary, within these boundaries, the quantitative results indicate that LLMs display (i) a reduced degree of syntactic variability compared to human-authored texts, and (ii) the range of variation remains more constrained than in authentic journalistic writing, whether adhering to SVO order or not.

6 Conclusion: How English influences AI's German output

The linguistic interference between English and German manifests most clearly in the preference for unmarked SVO word order, which appears in model outputs with an average frequency of 64.3% (ranging from 38.1% to 83.1%) compared to 45.5% in human-authored articles. Such figures point to a structural pressure toward English-like syntax, contributing to the observed *narrowing* of German sentence structure in model productions. However, the typological influence of English also emerges in the model's handling of the topological field, specifically regarding the rigidity of the sentence brackets. While the corpus records strict ungrammaticalities only occasionally, we observe a tendency towards premature field closure or unmotivated extraposition. Example (39) illustrates a case in a subordinate clause introduced by *was* 'what'.

(39) In diesem Artikel werden wir uns mit den Details des Pakets auseinandersetzen und analysieren, was *es bedeutet* für die Zukunft unseres Landes.
'In this article, we will address the details of the package and analyse what it means for the future of our country.' [L7]

Here, the finite verb *bedeutet* precedes the prepositional phrase. In standard German topology, subordinate clauses require Verb-Final placement, creating a brace that encloses the arguments. While the placement of the prepositional phrase (*für die Zukunft unseres Landes*) in the post-field is a syntactically permissible operation known as extraposition, often used to alleviate complex sentence structures (cf. Bech 1955) its usage here lacks the typical motivation of end-weight. The phrase is neither exceptionally long nor complex. Consequently, this linearisation strongly mirrors the English SVO pattern (*what it means for the future of our country*) rather than the standard German brace construction shown in (40).[22]

(40) […], was *es* für die Zukunft unseres Landes *bedeutet*.

Example (41) illustrates a similar phenomenon in a main clause, where the verbal complex *ist geprägt* 'is shaped' appears effectively adjacent, contrary to the standard principle which usually places the participle in the clause-final position.

(41) Die aktuelle politische Lage *ist geprägt* von Unsicherheit und Herausforderungen.
'The current political situation is marked by uncertainty and challenges.' [M20]

(42) Die aktuelle politische Lage *ist* von Unsicherheit und Herausforderungen *geprägt*.

While Example (41) is not ungrammatical (as post-verbal placement of PPs is common in spoken German and relaxed registers), it represents a stylistic deviation in the context of formal written news. The model essentially dissolves the sentence bracket entirely. This suggests a cross-linguistic interference where the model aligns the German participle *geprägt* immediately with the auxiliary *ist*, mimicking the continuous English structure 'is shaped by', rather than sustaining the suspension required by the German syntactic frame (cf. Example 42).

Such qualitative examples, though limited in number, reinforce the broader quantitative observation: AI-generated German is subtly but consistently shaped by

[22] It is telling that even standard commercial grammar checking algorithms flag the linearisation in Example (39) as a stylistic error, automatically suggesting the closed sentence bracket in (40) as the correct alternative. This reinforces the observation that while extraposition is theoretically permissible, its usage here is statistically and stylistically rejected by the codified norm.

English structural patterns. This interference does not stem from a model's inability to generate *native-like* German constructions; instead, such constructions appear with diminished frequency within the training dataset, thus indicating the necessary role of English as a pivot language in generating text in other languages. These evident patterns, therefore, direct attention to the imperative of critically examining the composition of training data, wherein English maintains an overwhelming dominance, thereby influencing the structural tendencies of model output in other languages. Consequently, the syntactic profile of AI-generated German cannot be regarded as a neutral replication of native newspaper syntactic patterns. Instead, it exhibits the typological imprint of English.

In summary, while certain features may meet equivalence thresholds for narrowly defined variables, these results must not be generalised as evidence that *LLMs write like humans*. Establishing such a claim would require a substantially broader analytical framework that integrates multiple syntactic, lexical, and discourse-level criteria, supplemented by further qualitative evaluation. Future research should expand the corpus, encompass additional text types, languages and models, and extend the investigation to lexical, textual, and stylistic dimensions in order to more fully capture the scope and limits of LLM language production. Specifically, research ought to address the core concern of training data composition and strive towards the development of more linguistically balanced datasets that more accurately represent languages (and language distribution) beyond English. Until such research is undertaken, the present findings serve as a reminder that the linguistic authenticity of AI-generated texts is inevitably influenced by the structural prevalence of English.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this contribution.

Acknowledgements

The research presented in this paper was conducted as part of a master's thesis in the joint double degree programme between the University of Trento and the TU Dresden, under the supervision of Professor Anna-Maria De Cesare Greenwald and Professor Ermenegildo Bidese. I owe them my sincere thanks for their mentorship, insightful feedback, and constant encouragement. Portions of this research were presented at the *Conversazioni Linguistiche* conference series at the University of Trento and at the 3rd international conference *Automated texts in the Romance and Germanic languages* (AI-ROM-III) at the TU Dresden.

Sources

Anthropic. https://claude.ai/ (last accessed on 18/08/2025).
BILD. https://www.bild.de/ (last accessed on 18/08/2025).
Mistral AI Team. https://mistral.ai/ (last accessed on 18/08/2025).
Nvidia. https://build.nvidia.com/meta/llama-3_1-405b-instruct (last accessed on 18/08/2025).
OpenAI. https://chatgpt.com/ (last accessed on 18/08/2025).
Süddeutsche Zeitung. https://www.sueddeutsche.de/ (last accessed on 18/08/2025).

SZ1: https://www.sueddeutsche.de/politik/steuerschaetzung-haushaltsstreit-ampel-koalition-scholz-lindner-habeck-industrie-lux.RzQBn1iW1K2d4xXMiigPx2
SZ2: https://www.sueddeutsche.de/meinung/globale-minderausgabe-haushalt-ampel-steuerschaetzung-lux.XFnze2fnfh3N6hT3AQDdWD
SZ3: https://www.sueddeutsche.de/politik/kranker-mann-europas-sozialbeitraege-lohnnebenkosten-lux.8apZu3D4eoj34pgRpM1hy3
SZ4: https://www.sueddeutsche.de/politik/baden-wuerttemberg-gruene-cem-oezdemir-winfried-kretschmann-manuel-hagel-landtagswahlen-lux.8REEn27LMKNgVwh1YyarQ4
SZ5: https://www.sueddeutsche.de/politik/thueringen-bsw-cdu-koalition-verhandlungen-lux.UXgB2FcPNao2iKoGaHL8hR
SZ6: https://www.sueddeutsche.de/politik/giuliani-anwalt-trump-gericht-schadenersatz-uhren-mercedes-lux.M7XHigxUr5kHqyA8yXhga5
SZ7: https://www.sueddeutsche.de/politik/bulgarien-wahl-prorussisch-apathie-nationalisten-regierung-lux.2DFuBEzDta5KEkrPbuQ9vU
SZ8: https://www.sueddeutsche.de/politik/eu-parlament-afd-haushalt-weber-evp-kritik-lux.L9aRhSnR4YnmdGNWET6grM
SZ9: https://www.sueddeutsche.de/politik/israel-hisbollah-libanon-bank-gold-geld-krankenhaus-lux.TxxwzAtVsoRuvgqkHfMiuy
SZ10: https://www.sueddeutsche.de/politik/japan-verbeugung-bedeutung-lux.BVR8MWCSYHRSRJenyF6P3T
SZ11: https://sueddeutsche.de/politik/bundesregierung-lindner-habeck-scholz-ampel-streit-lux.VxRSQofMX1vAgA5GkW8xaz
SZ12: https://www.sueddeutsche.de/politik/oezdemir-baden-wuerttemberg-ministerpraesident-lux.UdPdkd3wgQYCba6z5fzPcm
SZ13: https://www.sueddeutsche.de/politik/bundesregierung-streit-haushalt-luecke-lux.H8X3SCRHPYqDEaJjZ4jocU
SZ14: https://www.sueddeutsche.de/politik/franziskus-enzyklika-liebe-weltsynode-lux.CH55tg9wZ9nkAQtFP4epWE
SZ15: https://www.sueddeutsche.de/politik/polen-flucht-migration-asylpolitik-lux.9Ta9Jy5End5UwYo8wkYmio

SZ16: https://www.sueddeutsche.de/politik/doener-baden-wuerttemberg-heilbronn-lux.NA9U3BrsuNtfo2nKLPUSQo
SZ17: https://www.sueddeutsche.de/politik/steuer-bund-2028-fehlbetrag-58-milliarden-euro-lux.VrqjeB8U3QNRnVLtETsqMY
SZ18: https://www.sueddeutsche.de/politik/bundestag-antisemitismus-resolution-wissenschaft-documenta-lux.F15cGH8niLGEcyj4WiEFmS
SZ19: https://www.sueddeutsche.de/politik/tuerkei-terror-zurueck-pkk-regierung-luftangriffe-lux.2oYL6gkhzmhch5PnDw9pdw
SZ20: https://www.sueddeutsche.de/politik/ukraine-g7-kredit-christian-lindner-russland-trump-lux.YSTqBd6AcigqPbNESpKRYT
SZ21: https://www.sueddeutsche.de/politik/tuerkei-anschlag-ankara-erdogan-lux.QK9mF3K5beP7bJHnmVirZH
SZ22: https://www.sueddeutsche.de/politik/us-wahl-2024-trump-labour-demokraten-wahlkampf-beschwerde-lux.7DMBAmyrnTRYUyo8SrCTYh
SZ23: https://www.sueddeutsche.de/politik/deutschland-grossbritannien-verteidigungsbuendnis-russland-trump-lux.WEU3mWTZcEkUp44a2eowaT
SZ24: https://www.sueddeutsche.de/politik/healey-grossbritannien-verteidigungsminister-labour-lux.NtPrxDQ444bSvSjpKspR8y
SZ25: https://www.sueddeutsche.de/medien/liebe-angst-dokumentarfilm-eltern-kind-beziehung-traumata-sandra-prechtel-lux.X5ghEHe9oLMfUcbmgHqqsc

B1: https://www.bild.de/politik/ausland-und-internationales/us-wahl-hier-donnert-trump-im-muellwagen-gegen-kamala-harris-6722cba6b54ff317ce96b690
B2: https://www.bild.de/politik/inland/tote-hosen-staatspreis-wie-ein-dreckiger-sieg-im-fussball-672233408f9b8c158219fb6a
B3: https://www.bild.de/politik/inland/rundfunk-reform-gehalt-der-ard-bosse-soll-halbiert-werden-672219058f9b8c158219f958
B4: https://www.bild.de/politik/inland/naechster-ego-gipfel-scholz-laedt-lindner-und-habeck-wieder-nicht-ein-67222d33925a135303794e0f
B5: https://www.bild.de/politik/ausland-und-internationales/us-wahl-2024-muell-eklat-biden-patzer-koennte-harris-den-sieg-kosten-6722152bb54ff317ce96af3d
B6: https://www.bild.de/politik/ausland-und-internationales/islam-experte-rechnet-ab-multikulti-hat-die-clans-maechtig-gemacht-671f4cc9b7e14b57a92fa88a
B7: https://www.bild.de/politik/inland/fdp-macht-d-day-papier-oeffentlich-beginn-der-offenen-feldschlacht-67488a0afc23cc2c0725bba3
B8: https://www.bild.de/politik/medienbericht-bulgaren-gestehen-spionage-fuer-ex-wirecard-chef-marsalek-67487ed374f3603217752fa1

B9: https://www.bild.de/politik/inland/heiligabend-als-feiertag-in-polen-bald-realitaet-und-in-deutschland-67484ce877cfda51b7d0cf70
B10: https://www.bild.de/politik/inland/usa-die-gefahr-ist-real-fliehen-deutsche-firmen-jetzt-zu-trump-674731c4bf2bc404c5544e64
B11: https://www.bild.de/politik/inland/frauen-gesetz-gegen-gewalt-steht-vor-dem-aus-union-hat-andere-plaene-67447a9590aa3e14a1bd942d
B12: https://www.bild.de/leben-wissen/auto/lauschangriff-polizei-fordert-zweitschluessel-von-autoherstellern-6745d46a560fd42b3982060d
B13: https://www.bild.de/politik/ausland-und-internationales/usa-cowboy-lady-will-donald-trumps-erstes-abschiebe-lager-bauen-67474f3be5cd8c482209deeb
B14: https://www.bild.de/politik/ausland-und-internationales/syrien-rebellen-ueberrennen-assads-truppen-und-nutzen-ukraine-taktik-674833b397852e67ec3ecac7
B15: https://www.bild.de/politik/ausland-und-internationales/putins-bomben-millionen-ukrainer-ohne-strom-674827ec97852e67ec3eca3b
B16: https://www.bild.de/politik/ausland-und-internationales/grausame-rache-putins-massaker-an-ukrainischen-soldaten-67472d5f0b2db5219f08b270
B17: https://www.bild.de/politik/inland/herbst-der-entscheidung-zieht-einer-der-ampel-den-stecker-66f1245fc3ccdc3a464ca008
B18: https://www.bild.de/politik/inland/union-schlaegt-alarm-drei-not-massnahmen-sollen-auto-branche-retten-66f02454a6fc5611b0aaecb9
B19: https://www.bild.de/politik/ausland-und-internationales/first-lady-leitet-sitzung-werden-die-usa-von-jill-biden-regiert-66f17935e1d06e35845f06ff
B20: https://www.bild.de/politik/ausland-und-internationales/israel-attacke-erdogans-ruestungs-firma-wirbt-mit-eigenen-walkie-talkies-66f1283fc3ccdc3a464ca06a
B21: https://www.bild.de/politik/ausland-und-internationales/italien-hammer-keine-strafzettel-mehr-fuer-deutsche-66f10cc5dbe3804b995ee0b2
B22: https://www.bild.de/politik/ausland-und-internationales/aus-angst-vor-israel-aufstand-gegen-den-hisbollah-boss-66efeb1ec3ccdc3a464c92d7
B23: https://www.bild.de/politik/inland/programmzoff-in-der-cdu-wie-soll-friedrich-merz-deutschland-reparieren-66ed85c32d03d60867788589
B24: https://www.bild.de/politik/ausland-und-internationales/donald-trump-will-bei-wahlniederlage-nicht-mehr-antreten-keine-kandidatur-2028-66f0467fc3ccdc3a464c97d6
B25: https://www.bild.de/politik/ausland-und-internationales/hisbollah-unter-druck-israel-hat-terroristen-laecherlich-gemacht-66ef1ca3e1d06e35845ef11d

References

Backus, Ad & Cohen, Michael & Cohn, Neil & Faber, Myrthe & Krahmer, Emiel & Laparle, Schuyler & Maier, Emar et al. 2023. Minds: Big questions for linguistics in the age of AI. *Linguistics in the Netherlands* 40(1). 301–308. https:/doi.org/10.1075/avt.00094.bac

Bech, Gunnar. 1955. *Studien über das deutsche Verbum Infinitum*. København: I kommission hos E. Munksgaard.

Brown, Tom B. & Mann, Benjamin & Ryder, Nick & Subbiah, Melanie & Kaplan, Jared & Dhariwal, Prafulla & Neelakantan, Arvind et al. 2020. Language models are few-shot learners. *arXiv*. https:/doi.org/10.48550/arXiv.2005.14165

Cicero, Francesco. 2023. L'italiano delle intelligenze artificiali generative. *Italiano LinguaDue* 15(2). 733–761. https:/doi.org/10.54103/2037-3597/21990

Cicero, Francesco. 2025. Esercizi di stile. La narrativa per l'infanzia delle intelligenze artificiali. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 2(2). https:/doi.org/10.62408/ai-ling.v2i2.19

De Cesare, Anna-Maria. 2007. Le funzioni del passivo agentivo. Tra sintassi, semantica e testualità. *Vox Romanica* 66. 32–59.

De Cesare, Anna-Maria. 2023a. Giorgia Meloni, Meloni o la Meloni? La codifica degli antroponimi femminili in biografie generate da ChatGPT e pubblicate su Wikipedia. *Lingue e culture dei media* 7(1–2). 1–20. https:/doi.org/10.54103/2532-1803/22388

De Cesare, Anna-Maria. 2023b. Assessing the quality of ChatGPT's generated output in light of human-written texts: A corpus study based on textual parameters. *CHIMERA: Revista de Corpus de Lenguas Romances y Estudios Lingüísticos* 10. 179–210. https://revistas.uam.es/chimera/article/view/17979 (last accessed on 09/01/2026).

De Cesare, Anna-Maria. 2024. Nuove dinamiche di contatto linguistico. Le "impronte digitali" dell'inglese nell'italiano generato da LLM. *Lid'O. Lingua Italiana d'Oggi* XXI: 67–91.

De Cesare, Anna-Maria. 2025. LLM referential chain generation: A qualitative case study based on Italian biographies produced by GPT-4. *Linguistik Online* 136(4). 25–52. https:/doi.org/10.13092/8tppdv47

De Cesare, Anna-Maria & Garassino, Davide & Agar Marco, Rocío & Albom, Ana & Cimmino, Doriana. 2016. *Sintassi marcata dell'italiano dell'uso medio in prospettiva contrastiva con il francese, lo spagnolo, il tedesco e l'inglese*. Berlin: Peter Lang.

Deutscher Presserat. 2025. Publizistische Grundsätze (Pressekodex). Berlin: Deutscher Presserat.

Eberhard, David M. & Simons, Gary F. & Fennig, Charles D. 2025. *Ethnologue: Languages of the World*. 28th ed. Dallas: SIL International. https://www.ethnologue.com (last accessed on 21/12/2025).

Grattafiori, Aaron & Dubey, Abhimanyu & Jauhri, Abhinav & Pandey, Abhinav & Kadian, Abhishek & Al-Dahle, Ahmad & Letman, Aiesha et al. 2024. The Llama 3 herd of models. *arXiv*. https:/doi.org/10.48550/arXiv.2407.21783

Guo, Yanzhu & Conia, Simone & Zhou, Zelin & Li, Min & Potdar, Saloni & Xiao, Henry. 2025. Do Large Language Models have an English accent? Evaluating and improving the naturalness of multilingual LLMs. In Che, Wanxiang & Nabende, Joyce & Shutova, Ekaterina & Pilehvar, Mohammad Taher (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 3823–3838. Vienna: Association for Computational Linguistics. https:/doi.org/10.18653/v1/2025.acl-long.193

Hofstätter, Alexandra. 2020. *Entwicklungen in der deutschen Nachrichtensprache*. PhD. Dissertation. München: Ludwig-Maximilians-Universität München.

Holmberg, Anders. 2015. Verb Second. In Kiss, Tibor & Alexiadou, Artemis (eds), *Syntax – Theory and Analysis. An International Handbook. Volume 1*, 342–383. Berlin: De Gruyter Mouton. https:/doi.org/ 10.1515/9783110377408.342

Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e. V. (IVW). 2024. IVW - Quartalsausweisung. https://quartalsausweisung.ivw.de/ (last accessed on 18/08/2025).

Johnson, Rebecca L. & Pistilli, Giada & Menédez-González, Natalia & Duran, Leslye Denisse Dias & Panai, Enrico & Kalpokiene, Julija & Bertulfo, Donald Jay. 2022. The ghost in the machine has an American accent: value conflict in GPT-3. *arXiv*. https:/doi.org/10.48550/arXiv.2203.07785

Koch, Peter & Oesterreicher, Wulf. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36(1). 15–43. https:/doi.org/10.1515/9783110244922.15

Kurz, Josef & Müller, Daniel & Pötschke, Joachim & Pöttker, Horst & Gehr, Martin. 2010. *Stilistik für Journalisten*. Wiesbaden: VS Verlag für Sozialwissenschaften. https:/doi.org/10.1007/978-3-531-92492-2

Linden, Peter. 2000. *Wie Texte wirken: Anleitung zur Analyse journalistischer Sprache*. 2., aktualisierte und erw. Aufl. Berlin: ZV.

Linden, Peter. 2023. *Handbuch Stilsicher schreiben: Wie Sie wirkungsvolle Texte formulieren.* Berlin: Duden.

LiveBench. https://livebench.ai/#/ (last accessed on 18/08/2025).

LMArena. https://lmarena.ai/ (last accessed on 18/08/2025).

Lünenborg, Margreth. 2013. Boulevardisierung. In Bentele, Günter & Brosius, Hans-Bernd & Jarren, Otfried (Hrsg.). *Lexikon Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer. 33.

Matthies, Jochen. 2025. Agensdemotion und wissenschaftliche Verfasserreferenz im Kontext des KI-gestützten Schreibens – Eine Pilotstudie. In Lykhnenko, Olha & Matthies, Jochen & Roelcke, Thorsten & Teufele, Lisa (eds.), *Deutsch als Fremdsprache – Mehrsprachigkeit, Fachkommunikation, Digitalisierung und Kultur: Internationale Nachwuchstagung 2024 an der Technischen*

Universität Berlin, 91–114. Berlin, Heidelberg: Springer. https:/doi.org/10.1007/978-3-662-70351-9_5

McEnery, Tony & Hardie, Andrew. 2012. *Corpus linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Meyn, Hermann. 1985. *Massenmedien in der Bundesrepublik Deutschland*. Überarbeitete Neuauflage. Berlin: Colloquium Verlag.

Mittelberg, Ekkehart. 1967. *Wortschatz und Syntax der Bild-Zeitung*. Marburger Beiträge zur Germanistik. Marburg: N. G. Elwert Verlag.

Newman, Nic & Fletcher, Richard & Eddy, Kirsten & Robertson, Craig T. & Nielsen, Rasmus Kleis. 2023. *Reuters Institute Digital News Report 2023*. Oxford: Reuters Institute for the Study of Journalism.

Newman, Nic & Fletcher, Richard & Robertson, Craig T. & Ross Arguedas, Amy & Nielsen, Rasmus Kleis. 2024. *Reuters Institute Digital News Report 2024*. Oxford: Reuters Institute for the Study of Journalism.

Newman, Nic & Ross Arguedas, Amy & Robertson, Craig T. & Nielsen, Rasmus Kleis & Fletcher, Richard. 2025. *Reuters Institute Digital News Report 2025*. Oxford: Reuters Institute for the Study of Journalism.

OpenAI. 2023. OpenAI Data Partnerships. https://openai.com/index/data-partnerships/ (last accessed on 18/08/2025).

OpenAI. 2025. Best practices for prompt engineering with the OpenAI API. *OpenAI Help Center*. https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api? (last accessed on 18/08/2025).

Phoenix, James & Taylor, Mike. 2024. *Prompt Engineering for Generative AI*. O'Reilly Media, Inc.

Qin, Libo & Chen, Qiguang & Zhou, Yuhang & Chen, Zhi & Li, Yinghui & Liao, Lizi & Li, Min & Che, Wanxiang & Yu, Philip S. 2025. A survey of multilingual large language models. *Patterns* 6(1). 101118. https:/doi.org/10.1016/j.patter.2024.101118

Raabe, Johannes. 2013a. Boulevardpresse. In Bentele, Günter & Brosius, Hans-Bernd & Jarren, Otfried (Hrsg.). *Lexikon Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer. 33–34.

Raabe, Johannes. 2013b. Qualitätszeitungen. In Bentele, Günter & Brosius, Hans-Bernd & Jarren, Otfried (Hrsg.). *Lexikon Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer. 288.

Schmitt, Uta. 2004. *Diskurspragmatik und Syntax: die Funktionale Satzperspektive in der französischen und deutschen Tagespresse unter Berücksichtigung einzelsprachlicher, pressetyp- und textklassenabhängiger Spezifika*. Frankfurt am Main etc.: Peter Lang.

Shi, Freda & Suzgun, Mirac & Freitag, Markus & Wang, Xuezhi & Srivats, Suraj & Vosoughi, Soroush & Chung, Hyung Won et al. 2022. *Language Models are Multilingual Chain-of-thought Reasoners. ICLR 2023*.

SPIEGEL-Verlag (ed.). 2021. *Die SPIEGEL-Standards: Haltung, Recherche, Schreibweise*. Hamburg: SPIEGEL-Verlag.

Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1(1). https:/doi.org/10.62408/ai-ling.v1i1.14

Weinrich, Harald. 2007. *Textgrammatik der deutschen Sprache*. 4. Aufl. Hildesheim: Olms.

White, Colin & Dooley, Samuel & Roberts, Manley & Pal, Arka & Feuer, Ben & Jain, Siddhartha & Shwartz-Ziv, Ravid et al. 2025. LiveBench: A Challenging, Contamination-Limited LLM Benchmark. *arXiv*. https:/doi.org/10.48550/arXiv.2406.19314

Wöllstein, Angelika. 2022. *Duden: die Grammatik* (Der Duden in zwölf Bänden, Band 4). 10., völlig neu verfasste Auflage. Berlin: Dudenverlag.

Zheng, Lianmin & Chiang, Wei-Lin & Sheng, Ying & Li, Tianle & Zhuang, Siyuan & Wu, Zhanghao & Zhuang, Yonghao et al. 2024. LMSYS-Chat-1M: A large-scale real-world LLM conversation dataset. *arXiv*. https:/doi.org/10.48550/arXiv.2309.11998

Abbreviations

B: *BILD*
C: Anthropic's Claude 3.5 Sonnet
L: Meta's llama-3.1-405b-instruct
M: Mistral's mistral-large-2407
O: OpenAI's o1
SZ: *Süddeutsche Zeitung*

Appendix

Table A: Absolute figures for word order.

Syntactic Feature	SZ	BILD	o1	Claude 3.5	Llama 3.1	Mistral Large
+SVO	423	333	145	311	292	398
–SVO	643	274	236	219	64	112
non-sentential units	51	58	2	7	0	0
Σ	1,117	665	383	537	356	510