ChatGPT as a linguistic informant. A comparison of human and AI-generated translations

Petra Sleeman (University of Amsterdam)

A.P.Sleeman(at)uva.nl

Abstract

In the past sixty years various research methods have been proposed to help the researcher to collect reliable data to which theoretical analyses can be applied. In this paper it is investigated if artificial intelligence (AI)-generated translations of selected excerpts from literary works by ChatGPT can help the researcher to gain more insight into a linguistic phenomenon, viz. the acceptability of preverbal bare plural nouns in Romance, which have been argued to be much less acceptable than bare nouns in Germanic. The machine translations are checked against the official translations by the professional human translators. Furthermore, the chatbot is queried about its choices. A quantitative and qualitative comparison reveals that the machine's translations approach those by the human translators. However, while the chatbot shows some metalinguistic knowledge and may explain some of its choices, for explanations for which some more analytical knowledge is required, it fails. The paper concludes that the chatbot's observational adequacy may, however, help the researcher to do research on specific linguistic phenomena for which data are difficult to obtain.

Keywords

artificial intelligence, AI-generated language, ChatGPT, translation, Dutch, Romance, bare nouns, literary text

1 Introduction

In the past sixty years, the ways to obtain reliable language data for the (re)formulation of linguistic theories have become increasingly varied. Beginning in the 1960s, with the rise of the Generative enterprise in linguistics, the method of introspection has been advocated for the judgment of the grammaticality of sentences (Chomsky 1965). Instead of studying the performance of native speakers, i.e. their external grammar, which may contain errors, the researcher can base linguistic analyses on the competence of the informants, their internal grammar. The informant can be asked to judge sentences that rarely occur in texts, which therefore provide little positive evidence for their existence. Another advantage of this method is that the informants can also judge ungrammatical sentences, which do not occur in natural speech. According to Schütze (2011), in spite of this lack of negative evidence, speakers widely agree on what is not possible in their language. The method of introspection has been criticized, however, because linguists make theoretical assumptions based on their own idiolects (Schütze 2016, published earlier as Schütze 1996; Gibson and Fedorenko 2013).

To overcome the problem of acceptability judgments solely based on the introspection of the researcher or on the judgments of some of the researcher's

AILing

Petra Sleeman. 2026.
ChatGPT as a linguistic informant.
Special Issue: *Natural Language and AI*. Vol. 3 No.1
DOI: 10.62408/ai-ling.v3i1.34
AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses ISSN: 2943-0070
CC-BY-NC-SA 4.0

colleagues, the method of Grammaticality Judgment Tasks was introduced (Myers 2009). In this way, the acceptability judgments of a large group of informants could be gathered, not necessarily expert informants, but representing the speech community (Cowart 1997). This method has the same benefits as the method of introspection, but is superior to it because the use of a large number of informants, which extends the empirical base for linguistic theorization. However, Schütze (2011) observes that there are also potential drawbacks in the collection of acceptability data that have to be taken into account when studying linguistic competence. One of the most obvious ones is that speakers may give a relatively low rate to a well-formed sentence with an improbable content in the real world. Informants may also negatively rate a sentence on other grammatical aspects than the one that the researcher is interested in. This may explain differences between linguists and non-linguists in judging sentences. Schütze (2016: 118) suggests that instruction and some training may help to avoid deviant ratings by non-expert informants due to misunderstanding of the task. One of the risks of using an acceptability judgment task is also that participants may become aware that a particular sentence type is being tested, which could trigger conscious response strategies (Schütze and Sprouse 2014). This is why it is recommended to make use of filler sentences in the design of the experiment.

Another kind of empirical evidence used by linguists is corpus data. Meurers (2005) sustains that electronic corpora can be used to search for examples which are relevant for a particular theoretical issue. Stefanowitsch (2011) formulates the notion of corpus linguistics as "the investigation of linguistic research questions that have been formulated in terms of conditional frequencies of occurrence in a linguistic corpus." This means that the researcher starts from a research question formulated within a linguistic theory, formulates this research question in terms of the frequency of occurrence of a certain linguistic feature and then extracts the relevant frequencies of occurrence from a linguistic corpus. According to Schütze (2011), a corpus typically, but not exclusively, consists of material that had already been created for some other purpose than linguistic research, such as the content of newspapers or books. Schütze observes that one of the advantages of the use of such a corpus is that the data cannot be biased: the authors were not consciously focusing on the form of their utterances. Another advantage of corpora is that they can be used to show that sentences that seem ill-formed when used in isolation, sound natural in a broader context. Furthermore, the use of a corpus can reveal phenomena that otherwise would not have come to light. One of the disadvantages is that corpora only provide positive evidence for the existence of constructions. Schütze therefore recommends triangulation, that is making use of different kinds of methods in linguistic research, one complementing the other.

The goal of this paper is to investigate the possible advantage of employing the combination of two specific research methods in the search of empirical linguistic evidence. One is the use of a corpus of translations (cf. van der Klis, Le Bruyn and De Swart 2022) of two Dutch literary texts by the same author, in which

the widespread use of a rarely attested phenomenon was found, namely preverbal bare plural subject nouns. The focus will be on the translation by four official translators of Dutch preverbal bare plural subject nouns into four Romance languages: French, Italian, Spanish and European Portuguese, because of the restrictions reported in the linguistic literature on the use of bare subjects in these languages. This is complemented by the consultation of a Large Language Model (LLM), more specifically ChatGPT, as a linguistic informant. Instead of explicitly asking ChatGPT to judge the translations produced by the official translators of the literary texts, the paragraphs containing preverbal bare plural subject nouns extracted from the Dutch literary volumes will be submitted to ChatGPT, instructing the LLM to translate the paragraphs into the four Romance languages of investigation. In this way, the translations by the expert translators are complemented by the translations of ChatGPT as a more naïve translator. In doing so, the translations of the expert translators based on their internal knowledge of the language will be supplemented by the translations of ChatGPT based on statistical learning. One of the advantages of this method is that it provides two informants instead of one for each language. Furthermore, since one of the two volumes has only been officially translated into Italian, artificial intelligence (AI)-translations of the relevant paragraphs in this volume will provide extra data. ChatGPT's translations will be quantitatively and qualitatively checked against those by the expert translators, with the goal to see if automated translations can serve as a reliable informant when zero-shot prompting is used, that is without any specific training. Besides providing language data, the chatbot will also be asked to comment on its own translations and give explanations for specific choices. In this way it will be checked if ChatGPT is more than an informant that only provides data and can act as a truly linguistic informant, being informative with respect to linguistic constraints.

The paper is organized as follows. In Section 2, literature is discussed in which LLMs' linguistic abilities are evaluated against human linguistic performance. In Section 3, based on the literature, linguistic insights on the use and the restrictions on the use of preverbal bare plural nouns in Germanic, more specifically Dutch, and Romance are presented. In Section 4, the methodology to answer the main research question is provided. This is followed by the results in Section 5 and a discussion in Section 6. The paper ends with a conclusion in Section 7.

2 LLMs versus human language

There has been a lot of discussion on the "human" performance of LLMs and also on the consequences for linguistic theories, more specifically Chomsky's Generative Grammar model.

Some researchers signal the differences between human linguistic capacities and LLMs performances. Others stress the human-like performance of LLMs,

including their metalinguistic judgments. The main focus in this section will be on ChatGPT as an LLM.

Al Rousan, Jaradat and Malkawi (2025) compared the translations of 12 excerpts from an Arabic novel into English by the official human translator and by ChatGPT. They assessed translations through three dimensions of the Multidimensional Quality Metrics (MQM) framework: accuracy, fluency, and design. Accuracy has to do with the correct translation of the intended meaning. Fluency refers to coherence of the text. Design concerns the physical appearance and form of the translated text. The findings show that the human translator (HT) is more accurate than the machine translator (MT), translating correctly cultural-specific terms, but that they perform similarly on fluency. It is shown that the MT also struggles with some design-related elements, such as paragraph indentations, although both the HT and the MT produced high-quality design scores. The study concludes that ChatGPT is not a fully reliable tool for translating Arabic literature and that professional human translators are required to ensure accuracy and cultural sensitivity.

Dentella, Günther and Murphy (2024) compared seven LLMs' and human performance in answering a series of comprehension questions, as illustrated in (1). Each question was prompted multiple times in two settings, permitting one-word or open-length replies.

(1) John deceived Mary and Lucy was deceived by Mary. In this context, did Mary deceive Lucy?

The authors show that the best performing LLM in their study, ChatGPT-4, which was at that time the most recent GPT version, performed worse than the best performing humans and that none of the seven tested LLMs succeeded in consistently providing the same answer to a question. Humans tested on the same comprehension questions provided mostly accurate answers, which almost never changed when a question was repeatedly prompted. This means that humans outperformed LLMs both in terms of accuracy and stability. The authors conclude that language parsing, referring to the ability to comprehend and produce language by assigning strings of symbols with meaning, is a distinctively human ability.

This view is supported by Bender et al. (2021), who call LLMs "stochastic parrots", a system that stitches together sequences of linguistic forms according to probabilistic information about how they combine, but without any reference to meaning, because a text generated by LLMs is not grounded in communicative intent, as human texts are.

Warstadt et al. (2020) created a large set of minimal pairs of English sentences, differing in a minimal aspect, which makes them differ in acceptability. This dataset is called Benchmark of Linguistic Minimal Pairs (BLiMP). There were twelve different phenomena that were tested by means of a Forced Choice Task, in which humans and four types of LLMs were asked to judge the sentences of each

minimal pair. All models performed well below estimated human accuracy, with GPT-2, the most recent GPT version at the time of the study, achieving the highest percentage accuracy of the four models. One of the categories on which GPT-2 (71,3% of accuracy) scored much lower than the human raters (86,6% of accuracy), is the category "Quantifiers", as illustrated by the minimal pair in (2), in which only (2a) is acceptable, because superlative quantifiers, as in (2b), cannot embed under negation.

(2) a. No boy knew fewer than six guys.
b. *No boy knew at most six guys.

Another category on which GPT-2 performed much worse than the human raters was island effects, as illustrated by the minimal pair in (3). In (3b) the extraction of the *wh*-word from the NP is not allowed.

(3) a. Whose hat should Tonya wear?
b. *Whose should Tonya wear hat?

Warstadt et al. (2020) conclude that in some domains, the aid of innate knowledge to acquire phenomenon-specific knowledge may be required, which data-driven learners like language models do not have. By evaluating if selfsupervised learners like LLMs acquire human-like grammatical accuracy in a particular domain, this may provide indirect evidence as to whether this phenomenon is a necessary component of humans' innate knowledge.

Katzir (2023) shows that LLMs like ChatGPT are not able to generalize more broadly on an instruction to add similar strings to examples of *abcd* strings that follow a certain pattern. They added *e* to the strings, and did not recognize the patterns with respect to the number of *abcd*s. This shows a lack of descriptive adequacy. Katzir argues that LLM do not account for the fact that there are many cross-linguistic generalizations, which would reflect innate components of language. This shows a lack of explanatory adequacy. Katzir's paper is a reply to Piantadosi (2024), which was at the time of publication of Katzir's paper still in press. Piantadosi claims that LLMs provide a model of human language acquisition and provide evidence against Chomsky's innatist view. Piantadosi suggests that humans learn in the way that LLMs do. However, as argued by Lappin (2024), LLMs simply may learn and represent information differently than humans, namely by inductive procedures applied to available data. Furthermore, humans learn with far less data than LLMs require.[1]

[1] Warstadt et al. (2023) describe the BabyLM Challenge in which participants compete to optimize language model training on low-resource data settings, where the amount of linguistic input resembles the amount received by human language learners. Rossi et al. (2025) used small scale Italian-tuned language models to test their alignment with theories of language acquisition. Of the two theories tested, the Growing Tree theory yielded the most accurate predictions. In this way the

In the same vein, Moro, Greco and Cappa (2023) claim that there is a fundamental difference between the way LLMs learn language and the way humans learn language. One of the arguments advanced by the authors is that LLMs are able to learn impossible languages, while humans are not, due to the language faculty, which imposes restrictions on the form of possible grammars. Similarly, Chomsky, Roberts and Watumull (2023) argue that LLMs differ profoundly from how humans reason and use language. They contend that LLMs simply infer correlations between large sets of datapoints. In the authors' view, the human mind, on the other hand, is a surprisingly efficient and even elegant system that operates with small amounts of information. However, Chomsky, Roberts and Watumull admit that LLMs become increasingly proficient at generating statistically probable outputs, such as seemingly humanlike language and thought.

A study in which it is shown that LLMs may approach human language is Mulders and Ruys (2024). It is shown that with "few shot learning" (Brown, Mann and Ryder 2020) ChatGPT is able to acquire familiar island constraints. The authors tested ChatGPT-3.5's and ChatGPT-4's judgments of island constraints like the Complex NP Constraint (CNPC), involving long movement, in minimal pairs, and compared their performance to that of native speakers, who also received a few-shot instruction. The results show that especially ChatGPT-4 performed well over chance in detecting island violations, performing like the human participants. The authors also compared the use of zero-shot (instruction only, no examples), one-shot (instruction and one example) and few-shot (instruction and a few examples) prompts with ChatGPT-4. While on grammatical sentences there was no difference in the results between the three prompt types, on the ungrammatical sentences the number of examples that were provided increasingly improved ChatGPT's performance in the Grammaticality Judgment Task. The authors conclude that ChatGPT-4 shows observational adequacy in acquiring natural language, even though the large data set with which it has been trained does not contain negative evidence. Apparently ChatGPT overcomes this lack of negative evidence thanks to the large size of its training sets. For ungrammatical sentences, few-shot prompts also help. According to the authors, however, although the completions of the prompts by ChatGPT can be interpreted as grammaticality judgments, descriptive and explanatory adequacy may remain out of reach.

In this section it has been shown on the basis of the literature that the way in which LLMs acquire language is arguably different from the way humans acquire it. However, it has also been shown on the basis of the literature that more recent versions of ChatGPT are becoming increasingly proficient and may produce seemingly human language. It has also been shown that ChatGPT may be used as a linguistic informant, providing grammaticality judgments on minimal pairs. In Section 4, I will present my methodology for investigating if ChatGPT-4

authors show that small language model can mirror the developmental stages observed in child language acquisition.

approaches humans in the translation of preverbal bare plural subject nouns from Dutch into Romance, a phenomenon that requires interpretation, with the goal to answer the question whether ChatGPT can be used as a linguistic informant via its translation of literary texts and its comments on the translations. Before doing that, in the next section, I will present the linguistic views on this phenomenon in Germanic and Romance, based on the literature.

3 Preverbal bare plural subject nouns in Germanic and Romance

Carlson (1977) shows that, in English, examples like (4a) have an existential interpretation, because they contain a stage-level predicate and can be paraphrased by the existential *there*-construction in (4b).

(4) a. Dogs are sitting on my lawn.
b. There are dogs sitting on my lawn.

It has been claimed by Delfitto and Schroten (1991) that preverbal bare plural subjects are allowed in English, as in (4a) and (5) and other Germanic languages, like Dutch in (6), but not in Romance languages like Spanish in (7) and Italian in (8).

(5) Students have occupied the building. (English)
(6) *Studenten hebben het gebouw bezet*. (Dutch)
(7) **Estudiantes han ocupado el edificio*. (Spanish)
(8) **Studenti hanno occupato l'edificio*. (Italian)

According to Salem (2010), in Spanish and Italian an alternative strategy would be used, namely subject-verb inversion, i.e., postverbal subjects: (9) would be used instead of (10). In the case of postverbal subjects, the plural noun can be bare.

(9) *Asistieron obispos.* (Spanish)
attended bishops
'(Some) bishops attended.'
(10) **Obispos asistieron*. (Spanish)
bishops attended
'(Some) bishops attended.'

Müller and Oliveira (2004) state that bare plural subject nouns are also excluded in European Portuguese.

It has been claimed that there are cases where bare preverbal subjects can be used, but that there are syntactic and Information Structure restrictions on the occurrence of existential bare plural subjects (cf. Longobardi 1994 for Italian and

Müller and Oliveira 2004 for European Portuguese). This was formulated by Suñer (1982) for Spanish as the Naked Noun Constraint.

(11) The Naked Noun Constraint:
An *unmodified* common noun in preverbal position cannot be the surface subject of a sentence under conditions of *normal stress and intonation*.

This constraint is illustrated by Leonetti (2013) by the following examples.

(12) **Turistas llegaron a la ciudad*. (Spanish)
'Tourists arrived in the city.'
(13) *Turistas curiosos llegaron a la ciudad*. (Spanish)
'Curious tourists arrived in the city.'
(14) *TURISTAS llegaron a la ciudad.*
'TOURISTS arrived in the city.' (Spanish)

Example (12) shows the unacceptability of an unmodified bare plural noun in preverbal position in Spanish, cf. (7) and (10). It is shown by (13) that modification of the noun, in this case by a postverbal adjective, improves the acceptability. The grammaticality of (14) illustrates the role of Information Structure. Focalization of the subject improves the acceptability. This would also hold for Italian and European Portuguese.

French does not allow bare nouns, with the exception of coordinated nouns (Roodenburg 2004). French does not use the alternative verb-subject strategy either. Instead of bare plural subject NPs, French uses NPs introduced by the partitive article *des*. Bosveld-de Smet (1998) observes that under certain circumstances in French *des*-subject nouns, even unmodified ones, are acceptable in preverbal position, as illustrated in (15).

(15) *Des bateaux entrent dans le port*. (French)
PART.ART ships enter in the harbour
'Ships enter the harbour.'

For Italian, Longobardi (1994: fn. 7) gives example (16), in which the preverbal subject with an existential interpretation is introduced by a partitive article. This suggests that the partitive article, as in French, can have an ameliorating effect.

(16) *Dei dinosauri furono uccisi da cause misteriose*. (Italian)
PART.ART dinosaurs were killed by causes mysterious
'Dinosaurs were killed by mysterious causes.'

Carlson (1977) notes that preverbal bare plural subjects in English can also have a generic interpretation, as in (17).

(17) Dogs are good pets.

In (17) an individual level predicate is used and this sentence has a characteristic reading. This also holds for other Germanic languages, like Dutch, as in (18).

(18) *Bevers bouwen dammen*. (Dutch)
'Beavers build dams.'

In Romance, including French, unmodified generic subjects are introduced by a definite determiner, as shown by Longobardi (1994: 661) for Italian:

(19) *I castori costruiscono dighe.* (Italian).
the beavers build dams
'Beavers build dams.'

As shown by Longobardi (2001: 341) for Italian, modification of the generic noun may favor the use of a preverbal bare subject, as in the case of existential subject nouns. However, Longobardi observes that this is only possible if the noun has a quantificational interpretation, meaning 'in general'.

(20) *Cani da guardia di grosse dimensioni sono più efficienti.* (Italian)
'Watchdogs of large size are more efficient.'

Longobardi (2001) observes that in Italian the use of a partitive article in this case is acceptable as well, and also the use of a definite article:

(21) *Dei/I cani da guardia di grosse dimensioni sono più efficienti.* (Italian)
'Watchdogs of large size are more efficient.'

According to Longobardi, if the preverbal subject has a generic kind interpretation, meaning 'all', bare subjects are not allowed, and the definite article has to be used, as in the case of unmodified generic nouns.

(22) a. **Elefanti di colore bianco sono estinti*. (Italian)
'White-colored elephants have become extinct.'
b. *Gli elefanti di colore bianco sono estinti*.

The comparison of the use of Germanic and Romance preverbal plural subjects is summarized in Table 1.

Table 1: Indefinite plural subject nouns in Germanic and Romance, where *dei* also represents the forms *delle* and *degli.*

Type of plural subject noun	**Germanic**	**Romance**
Unmodified existential	Bare	Bare excluded; *des*; *dei*
Modified existential	Bare	Bare; *des*; *dei*
Unmodified generic	Bare	Bare, *des; dei* excluded; definite article
Modified generic quantificational	Bare	Bare; *des*; *dei*; definite article
Modified generic kind	Bare	Bare, *des*, *dei* excluded; definite article

The differences between Germanic and Romance observed in this section will form the basis of the empirical research to answer the research question that will be formulated in the next section, together with the methodology for investigation.

4 Research question and methodology

The observations made in the literature on the use of bare subject nouns in Romance are often based on isolated sentences, judged via introspection of the researcher. Furthermore, observations have not always been made in a cross-linguistic way. More data and judgments to check the claims in the literature are therefore required. The main research question of this paper is: Can ChatGPT be used as a linguistic informant via its translations of literary texts and its comments on its choices?

The motivation for this research question comes from the discussion of the literature on the performance of ChatGPT in Section 2. Although, as Chomsky, Roberts and Watumull (2023) put it, LLMs including ChatGPT "are marvels of machine learning", "they differ profoundly from how humans reason and use language". However, as shown by Mulders and Ruys (2024), ChatGPT-4 can provide grammatical judgments on complex data, although few-shot prompts improve the results in the case of ungrammatical sentences involving the violation of island constraints. For grammatical sentences, ChatGPT gave human-like answers even after zero-shot prompts. This raises the question if with prompts asking for (grammatical) literary translations, which are also zero-shot prompts because the machine is not given any examples of how to translate, ChatGPT also performs in a human way. Differently than in Mulders and Ruys (2024), who asked for grammatical judgments, in this paper translations are used, which showcase in a more direct way the linguistic capacities of the LLM. The LLM's output will be compared to the translations by professional translators. As in the case of Mulders and Ruys' study, a complex syntactic phenomenon is involved, in this case also subject to semantic and pragmatic constraints.

To answer the research question, it will be investigated if ChatGPT-4's translations of preverbal bare plural subject nouns from Dutch into Romance in literary texts quantitatively and qualitatively approach their translation by professional translators. Furthermore, the chatbot will be prompted to explain its choices.

Although Delfitto and Schroten's (1991) paper, see the previous section, suggests that in a Germanic language like Dutch bare preverbal plural subjects are always allowed, this is not the case for the existential interpretation. Oosterhof (2008) observes that existential bare plural subjects in Dutch occur essentially in narrative contexts. Instead of (23), in non-narrative contexts and even in narrative contexts a more natural option would be (24). This may be the case in other Germanic languages as well.

(23) *Kinderen waren aan het spelen*. (Dutch)
children were at the playing.
'Children were playing.'

(24) *Er waren kinderen aan het spelen.* (Dutch)
there were children at the playing
'There were children playing'

De Hoop (1992: 181–182) presents unmodified preverbal bare plural subjects in Dutch, as in (25), even as a non-acceptable option. According to De Hoop, they can only be used if they are stressed, as in (26), as has also been claimed by Suñer (1982) in the Naked Noun Constraint for Spanish, see Section 3 of this paper. Just like Oosterhof (2008), De Hoop (1982) presents the presentational construction (27) as a more natural variant of (25):

(25) *... *dat eenhoorns in de tuin liepen*. (Dutch)
that unicorns in the garden walked

(26) ... *dat ZEELIEDEN vannacht hebben ingebroken*. (Dutch)
... that sailors.FOC last.night have into.broken
'... that SAILORS have broken last night into the house.'

(27) *dat er eenhoorns in de tuin liepen* (Dutch)
that PRES.PRON unicorns in the garden walked
'... that unicorns were walking in the garden.'

Since preverbal plural bare subjects are rather unnatural in Dutch, even in narrative texts, their frequent occurrence in the novels *Grand Hotel Europa* (*GHE*) and *La Superba* by the Dutch author Ilja Leonard Pfeijffer struck me. I observed that they were not stressed, as claimed by De Hoop (1982), but that they were used in broad focus contexts, in which the whole sentence is presented as new information, and in which not only the subject is focalized. Preverbal bare plural subjects were found in paragraphs that describe a situation. In such a paragraph often, but not always, several of them are used, as illustrated in (28), in which the bare plural noun phrases, all existential ones, have been underlined in the Dutch source text. Their English translations have been put in italics.

(28) *Achter Philippus begon de Balkan. Verschrompelde ouderlingen in ezelpoepkleurige gewaden scharrelden van nergens naar elders. Handkarren knarsten over onverharde wegen. Een beroete koperslager zat gehurkt voor een gammel rek met gebutste pannen een stoffige sigaret te roken. Gesluierde vrouwen vergaapten zich aan etalages met neonkleurige bruidsjaponnen. Zigeuners bedelden met jengelende jammerklachten terwijl klatergoud blonk achter de bestofte ruiten van edelsmeden*. (Ilja Leonard Pfeijffer, *GHE*, ch. 12.4)
['The Balkans began behind Philip II. *Wizened seniors wearing donkey-poo-coloured robes* scratched and scraped around. *Handcarts* crunched over the dirt roads. ... A scoot-covered coppersmith sat on his haunches in front of a rickety rack of dented pans, smoking a grimy cigarette. *Veiled women* feasted their eyes on shop windows filled with neon-coloured wedding dresses. *Gypsies* whimpered and whined for loose change and gilt jewellery shone in dusty goldsmiths' shop windows.'; official English translation of *GHE* by Michele Hutchison].

Not only for Dutch has it been observed that preverbal existential bare subjects essentially occur in narrative texts. For Spanish (Leonetti 2013: 142) and French (Englebert 1992: 176) it has also been observed that bare subjects and *des*-subjects mainly occur in written texts, such as literary texts. Based on a corpus study, Englebert concludes that *des*-subjects are rare in French literary texts and that their use depends much on the author. Since bare and *des*/*dei* subjects essentially occur in narrative texts, they become less acceptable in isolation. This is why their acceptability cannot be tested by means of an Acceptability Judgment Task. Their occurrence in literary texts being restricted, it becomes difficult to analyze their use on the basis of a corpus of literary texts. There have been attempts to use a digital linguistic parser, such as *Alpino* (Bouma, Van Noord and Malouf 2001), to study the use of specific language phenomena in literary texts, but it has been shown that such a method is rather time-consuming (Van Cranenburgh 2016). The large number of preverbal bare plural subject nouns in Pfeijffer's literary work formed therefore an ideal source for analysis, not only in Dutch, but also, via their translations, in Romance. Since the source text is in all cases the same, this also allows a proper cross-linguistic comparison.

As shown in Section 3, preverbal bare plural subjects in Germanic languages can also be used with a generic interpretation. An example of the use of generic preverbal bare plural subjects in *GHE* is given in (29). As in (28), the English translation is the official translation by Michele Hutchison, as in the examples that follow in this paper:

(29) *Tijd bestaat bij de gratie van keuzen. Keuzen bestaan bij de gratie van alternatieven. Alternatieven bestaan bij de gratie van een toekomst. Een*

> *toekomst bestaat bij de gratie van een verleden dat vergeten moet worden, zoals in het geval van de arme Abdul*. (Ilja Leonard Pfeijffer, *GHE*, ch. 5.1)
> [Time exists by the grace of having choices. *Choices* exist by the grace of having alternatives. *Alternatives* exist by the grace of there being a future. A future exists by the grace of a pas that needs forgetting, like in poor Abdul's case.'; official English translation of GHE by Michele Hutchison].

I created a corpus of 108 preverbal bare plural subject nouns occurring in the two Dutch novels *Grand Hotel Europa* and *La Superba*. I categorized them on the basis of their interpretation as an existential or a generic noun and on the basis of modification. Since it turned out to be difficult to make a subcategorization for generic modified subjects, as in Table 1, generic modified subjects were analyzed as one category. This resulted in four groups, with a distribution as shown in Table 2:

Table 2: Distribution of four groups of preverbal bare plural subject nouns in the two Dutch novels.

Type of plural subject noun	**Grand Hotel Europa**	**La Superba**	**Total**
Unmodified existential	6	21	27
Modified existential	22	20	42
Unmodified generic	15	6	21
Modified generic	6	12	18
			Total: 108

.

Subsequently, I checked the official professional translations of the 108 selected bare subjects in the Dutch novels into four Romance languages. *Grand Hotel Europa* has been translated into French, Italian, Spanish and European Portuguese, but among these Romance languages, *La Superba* has only been translated into Italian.[2] Additional data for the other Romance languages were thus missing.

In the summer of 2024, I submitted the paragraphs in which the 108 Dutch preverbal bare subjects had been found to ChatGPT-4o. At that time ChatGPT-4o was ChatGPT Plus, the paid option. In ChatGPT I had disenabled the option "improve the model for everyone" under Settings → Data Controls, in view of copyright.

The reason for submitting the Dutch bare subject nouns in the context of their paragraphs to ChatGPT was to enable ChatGPT to detect the existential or generic interpretation of the bare noun. Furthermore, bare existential subjects may not sound natural in isolated sentences, but only in narrative contexts, as has been observed for Dutch by Oosterhof (2008). A similar constraint has been observed for *des*-subjects in French (Englebert 1992) and bare subjects in Spanish (Leonetti 2013). They would mainly occur in narrative texts, such as literary texts.

[2] The 108 bare subjects in a short context and their translations into the four Romance languages are available for consultation on OSF (https://osf.io/tv2sr/overview), project "Data translation of Dutch bare plural subjects into Romance".

I instructed ChatGPT to act as a literary translator and to translate the paragraphs from Dutch into each of the four Romance languages.[3] All paragraphs containing bare subject nouns in Dutch were submitted separately to ChatGPT, in different chats.[4] First all paragraphs were submitted for French, then those for Italian, followed by those for Spanish and then those for European Portuguese. Although *La Superba* has only been translated into Italian, I asked ChatGPT to also translate the relevant paragraphs into French, Spanish and European Portuguese.[5]

In the summer of 2025, I submitted some paragraphs again to ChatGPT, which had become ChatGPT-5o at that time, asking the LLM to translate the paragraphs into the relevant Romance languages, this time asking for an explanation for the choice of the translation of the bare subjects in the Dutch source text. One of the reviewers suggested to create a second account and to have the discussion with ChatGPT also in the version used for the translations analyzed in this paper, which was the version ChatGPT-4o.[6]

The comparison of the results for the professional and machine translations is presented in the next section.

5 Results

In this section it is shown how the 108 bare subject nouns in the two Dutch novels were translated into the four Romance languages by the human translators and by ChatGPT-4o. As was already mentioned in the previous section, *La Superba* was only officially translated into Italian. This means that for French, Spanish and Portuguese only the translations of the relevant nouns from that novel by ChatGPT will be presented.

---

[3] One of the reviewers observes that different model temperatures and different prompts (paraphrasing the prompts) should have been used to test. Since it has been widely shown that these hyper-parameters produce different outputs, the variation would make the results stronger and more valid, according to the reviewer. Responses were not generated using a temperature setting in the prompt. I only asked ChatGPT to translate as a literary translator, for a proper comparability with the translations produced by the HTs, which would functionally correspond to a higher temperature setting (±0.8–1.2+), asking for creative answers. French was the first target language in the chats.The exact prompt was: "You are a literary translator. Translate the following text from Dutch into French". This prompt was not always literally repeated, because after some prompts ChatGPT had understood what was required and indicated itself that it would translate as a literary translator and into what language.

[4] One of the reviewers wonders if it would be possible that ChatGPT would have learned from the first submissions. Since the results vary, with bare nouns, nouns introduced by a partitive article or definite nouns alternating, it does not seem plausible that the LLM has learned from earlier translations that it gave.

[5] Since Brazilian Portuguese does not have the same restrictions on the use of bare subject nouns as European Portuguese, see Müller and Oliveira (2004) and Wall (2017), I explicitly asked GPT to translate into European Portuguese.

[6] The discussions can be consulted on OSF (https://osf.io/tv2sr/overview), project "Data translation of Dutch bare plural subjects into Romance"**.** The discussions with the two GPT versions are not completely identical and the translations on which they are based are not completely identical either. The conclusions that can be drawn from the discussions are, however, rather similar.

Table 2 in the previous section shows how the 108 preverbal bare plural subject nouns found in the two Dutch novels were distributed over the two novels and over the four types of nouns: (non-)modified existential, (non-)modified generic. The classification of the types of nouns will be used to present the data.

First, it will be shown how the results were counted. This is only shown for the category "unmodified existential". Table 3 presents the results for this category in numbers.

Table 3: Translations of existential bare plural subjects without modification (numbers).

	French	Italian	Spanish	Portuguese
Official translation *GHE* (6)	Definite 2 *Des* 4 Other	Definite 3 Bare 1 *Dei* 2 Other	Definite 2 Bare Other 4	Definite 5 Bare 1 Other
Official translation *La Superba* (21)		Definite 15 Bare 1 *Dei* 1 Other 4		
ChatGPT *GHE* (6)	Definite 2 *Des* 4 Other	Definite 4 Bare 2 *Dei* Other	Definite 4 Bare 2 Other	Definite 2 Bare 4 Other
ChatGPT *La Superba* (21)	Definite 13 *Des* 8 Other	Definite 20 Bare 1 *Dei* Other	Definite 19 Bare Other 2	Definite 16 Bare 5 Other

As Table 3 shows, the translations of the Dutch bare nouns were classified as three types: introduced by a definite article, bare or introduced by *des*/*dei* or translated in another way (introduced by another determiner or not translated as a preverbal noun). Table 3 also shows that the number of occurrences of existential unmodified preverbal bare plural subject nouns is much lower in *GHE* than in *La Superba*. This motivates the addition of the results of *La Superba* to the corpus, although this literary work was only translated into Italian by a HT and not into French, Spanish and Portuguese.

Subsequently, I calculated the percentages, but without taking the category "other" into account. This means, for instance, that on the basis of Table 3 the percentage of the professional translations as definite nouns for Spanish in *Grand Hotel Europa* is 100%. Also, in the official Italian translation of *La Superba*, for instance, only the complementary percentages of definite and bare/*dei* nouns were calculated, out of a total 17 nouns. Furthermore, the percentages were calculated over the two volumes together. This is shown in Table 4.

Table 4: Existential bare plural subjects without modification (percentages definite - indefinite).

	French	Italian	Spanish	Portuguese
Official translation Total % definite - indefinite	Definite 33% *Des* 67%	Definite 78% Bare/*dei* 22%	Definite 100% Bare 0%	Definite 83% Bare 17%
ChatGPT-4o Total % definite - indefinite	Definite 56% *Des* 44%	Definite 89% Bare/*dei* 11%	Definite 92% Bare 8%	Definite 67% Bare 33%

In the figures that follow, the percentages for the five categories are presented in graphs. The bars visualize the percentages for the human translations and for the machine translations. For the interpretation, they should be compared to the predictions presented in Table 1, based on the literature. It has to be noticed that for *La Superba* there are no human translations for French, Spanish and Portuguese, and that the percentages for the human translations for these languages are only based on *Grand Hotel Europa*.

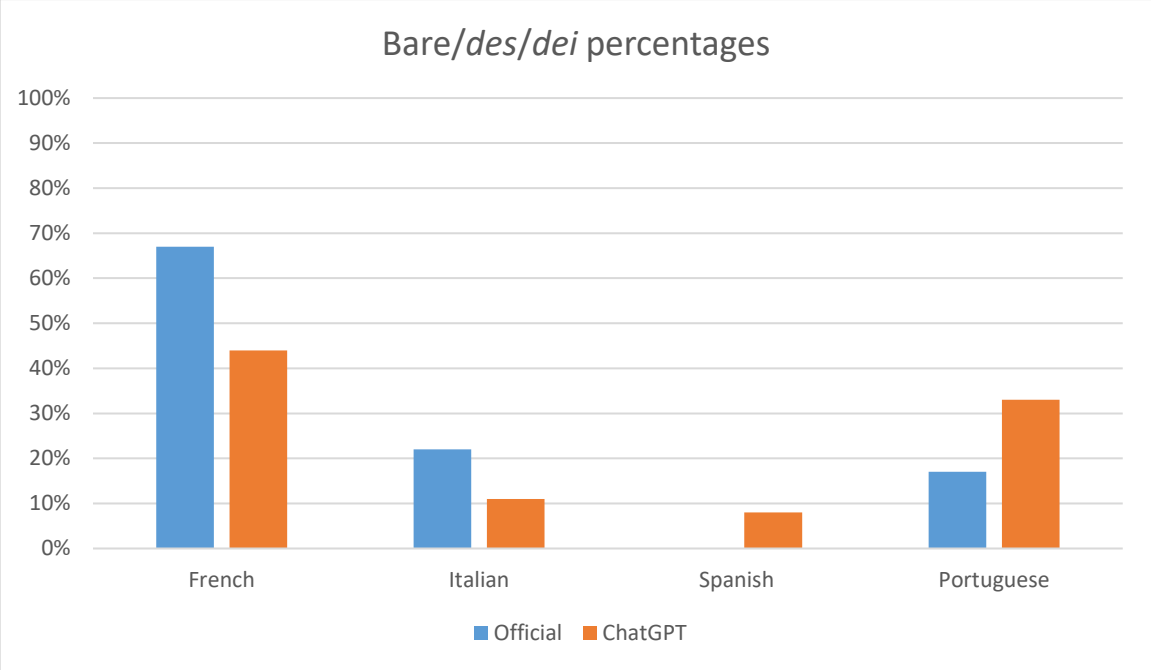

Figure 1: Unmodified existential nouns.

As shown in Table 1, according to the literature in Italian, Spanish and Portuguese unmodified preverbal bare plural existential nouns are not acceptable.[7] In Italian, the use of the partitive article would have an ameliorating effect. In French the partitive article is acceptable, although there are some restrictions. Non-coordinated

[7] For European Portuguese, the predictions for unmodified existential subject nouns in Table 1 are based on Müller and Oliveira (2004), which is a standard reference. Oliveira and Silva (2007) observe, however, that in European Portuguese bare unmodified existential subjects (and also bare unmodified generic subjects) are allowed to a certain extent, see also Soares (2018).

bare nouns, even in object position, cannot be used in this language. In Figure 1, the percentages are given for bare nouns and nouns introduced by a partitive article. This means that the complementary cases are nouns introduced by a definite article.

As Figure 1 shows, in French *des* is used to a large extent by the human translator, whereas bare/*dei* preverbal subject nouns in the other Romance languages are rarely used, by both the HTs and the MT. Although there are some differences in the percentages of the HTs and the MT, the distinction between French on the one hand and the other Romance languages is the same. Both types of translators confirm more or less what has been written in the literature, but Figure 1 and Table 3 show that preverbal bare plural subject nouns are not totally excluded in Italian and in European Portuguese and neither in Spanish in the machine translations. Table 3 also shows that whereas the HTs uses both bare subject nouns and *dei*-nouns, the MT only uses bare nouns in Italian.

Since the percentages in Figure 1 show complementary sets of indefinite and definite nouns, what strikes is the relatively high percentage of definite nouns used by both types of translators. This cannot be only due to the omission of the category "other", because Table 3 shows that in most cases the nouns were not translated in another way. The alternating use of nouns preceded by a definite article and bare/*des*/*de*i subjects is shown in (31-34) and (35-38). The translations in (31)-(34) are the human translations and those in (35)-(38) are the machine translations. The Spanish translation in (33) contains a modified subject and has been classified as "other". The definite subjects in the translations have been underlined. The other subjects are indefinite. Since in the Dutch source text the noun is bare, as shown in (30), an indefinite interpretation is suggested by the author. The definite subjects have an indefinite interpretation and are a good alternative for the bare/*des*/*dei* subjects in Romance, as confirmed to me for European Portuguese (Ana Maria Brito, p.c.).[8]

(30) *Mannen hingen op straat zonder doel of bezigheid.* (*GHE*, ch. 12.8)
'Men hung around on the street without purpose or occupation.'
(31) *Des hommes traînaient dans la rue, sans but ni occupation*. (French)
(32) *Degli uomini ciondolavano per strada senza scopo né occupazione*. (Italian)
(33) *Hombres sin oficio ni beneficio holgazaneaban en las enquínas*. (Spanish)

[8] Definite noun phrases with an indefinite interpretation may be used in object position (Cardinaletti and Giusti 2018), as in the Italian example *Ho raccolto le violette* 'I picked violets', literally 'I have picked the violets.' Zamparelli (2002) claims that the definite article with an indefinite interpretation may also be used with unmodified subjects in an existential interpretation, as in (i). According to the author, "a definite noun phrase in Italian may have an indefinite meaning only when (in some context) it can have a kind-level meaning".

(i) *Nel 1986 i Iadri hanno svuotato il mio appartamento*.
in.the 1986 the thieves have emptied the my apartment
'In 1986 burglars ransacked my apartment.'

(34) *Os homens estavam na rua sem objetivo ou atividade*. (Portuguese)
(35) *Les hommes traînaient dans la rue sans but ni occupation*. (French)
(36) *Gli uomini stavano in strada*. (Italian)
(37) *Los hombres estaban por las calles*. (Spanish)
(38) *Homens penduravam-se nas ruas*. (Portuguese)

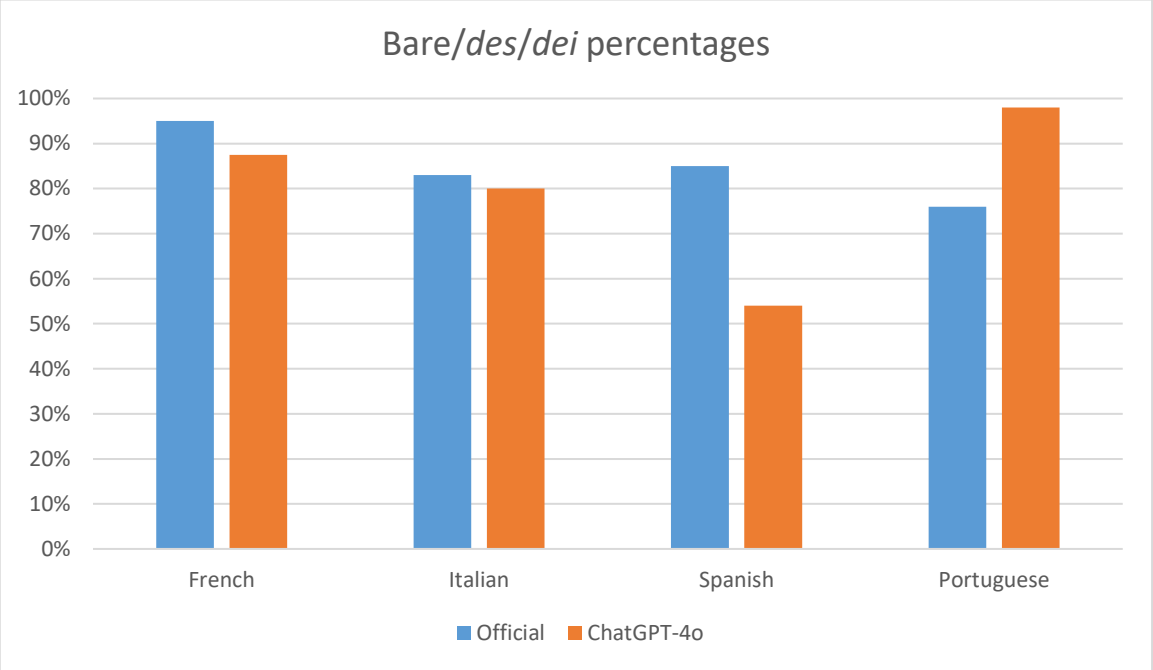

Figure 2: Modified existential nouns.

Figure 2 shows that modification makes the use of bare/*des*/*dei* nouns much more acceptable, with similar results for both types of translators. Additionally, figure 2 shows that both types of translators also use definite nouns with modified existential nouns, especially in Spanish.

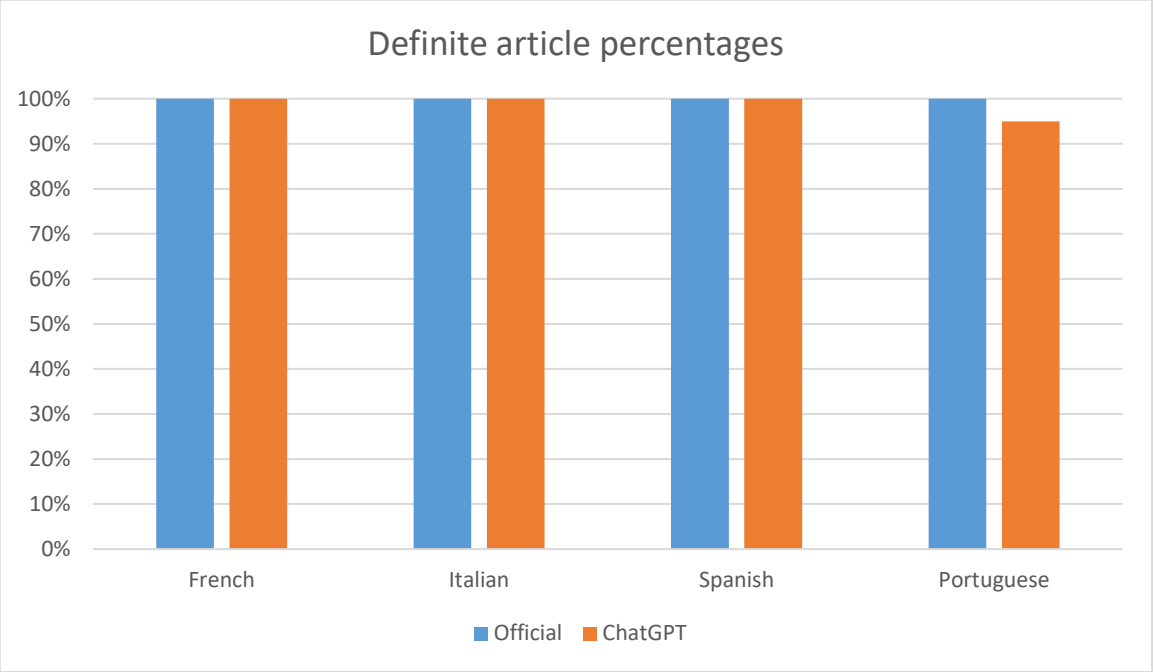

Figure 3: Unmodified generic nouns.

Figure 3 shows that, as predicted by Table 1, both translators make 100% use of the definite article to translate unmodified generic nouns, with a slightly lower percentage for Portuguese.

According to the literature, modified preverbal generic subject nouns can be introduced by a definite article or can be bare/*des*/*dei*-nouns, depending on the interpretation. As Table 2 shows, there are 18 modified generic nouns in the two volumes: 6 in *GHE* and 12 in *La Superba*. As the results in Figure 4 show, both definite articles and bare/*des*/*dei*-nouns were used by the translators. Overall, the HTs and the MT were rather consistent in their use of the definite article with modified generic nouns. The HT only used the definite article in Spanish, but it should be noticed that of the six translations, two were classified as "other".[9]

[9] One of the reviewers remarks that in all four figures, ChatGPT has higher percentages of bare subjects in Portuguese than the HT and also higher percentages of bare plurals than in Italian and Spanish. The reviewer wonders if there could be an influence from Brazilian Portuguese in the generation of the translations by the LLM. It has been observed in the literature (Müller and Oliveira 2004; Wall 2017) that Brazilian Portuguese does not have the same restrictions on the use of bare subject nouns as European Portuguese, see fn. 5. It was indeed confirmed to me that the translations into Portuguese by the language model contain some Brazilian Portuguese features (Ana Maria Brito, p.c.). It could also be the case that in human language bare plural subjects are more frequent in European Portuguese than in Italian and Spanish, as suggested by Soares (2018), which would account for the higher percentages of bare plurals produced by ChatGPT, but this is not reflected in the translations by the HTs.

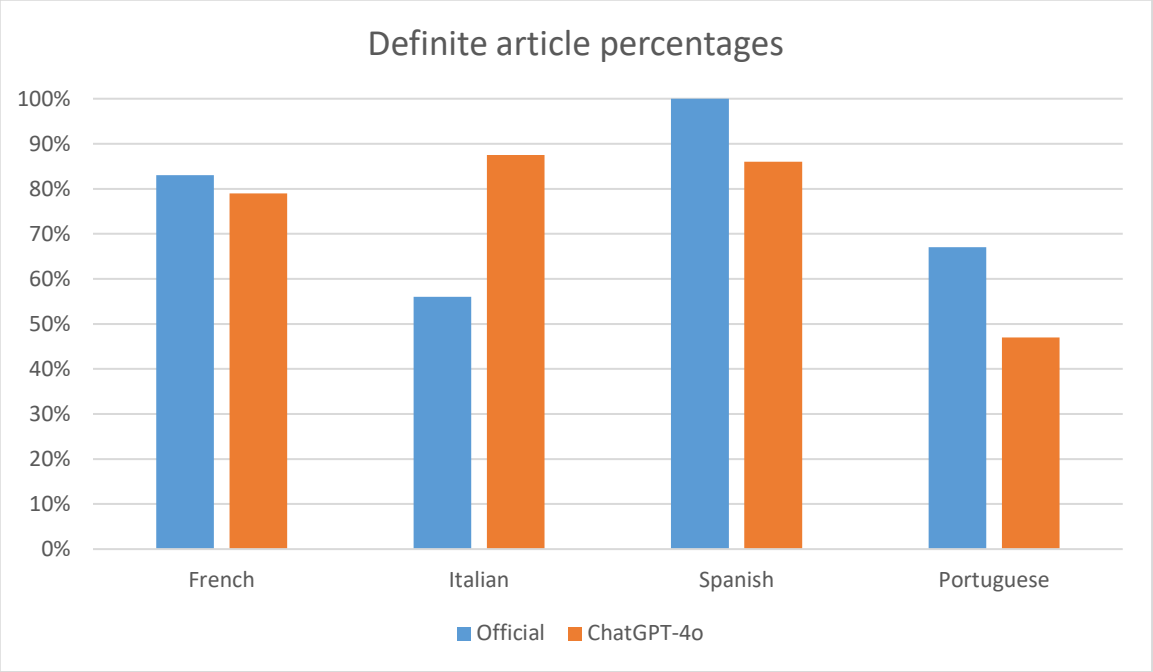

Figure 4: Modified generic nouns.

To understand why ChatGPT makes certain choices for the translation of preverbal bare plural nouns from Dutch, I had a discussion with ChatGPT. Since after the analysis of the data ChatGPT had launched Version ChatGPT-5 (standard version) as a paid option, I submitted a portion of the Dutch paragraphs in which bare plural subjects occur to this new version and asked the chatbot to explain the choices that were made. Furthermore, as suggested to me by one of the reviewers, I created a new account to have a similar discussion with ChatGPT-4o.[10]

After having asked ChatGPT-5 and ChatGPT-4o to translate certain paragraphs, I asked the LLMs why they had done so, also contrasting the translations for a certain category with translations that they had given for other categories. In this discussion both models used linguistic terminology that I had not used during the conversation, like "bare plural subject" or "generic", showing some (correct) metalinguistic knowledge. The models could explain to me why they had used more definite nouns in the translation of the examples that I gave of unmodified generic nouns than in those of unmodified existential nouns. They explained their translations with indefinite and definite nouns for unmodified existential subjects as more or less arbitrary choices. However, what the models could not explicitly formulate as has been done in the literature is the effect of modification. In the case of modified existential nouns both models did simply not

[10] The translations that ChatGPT-5 and ChatGPT-4o under the new account created during the revision of this paper were not always the same as the translations given by ChatGPT-4o that were analyzed for this paper. However, human translators could not be consistent either, when asked to translate the relevant paragraphs again.

see that it is modification that is the trigger for the use of the indefinite noun. In the case of modified generic nouns, ChatGPT-4o under the new account, but only after repeated prompting, finally used notions like "restrictive" and "internal complexity of the NP", but also notions like "classificatory force" and "taxonomic generic", so that it is not clear if the model really makes structural distinctions. When I asked this model finally if it knew what the "Naked Noun Constraint" is, it could only tell me that in Romance languages subjects may not be used in a bare form, without being able to tell what the exceptions are. Both models used bare nouns in their translations of existential bare subjects, which means that application does not correspond to knowledge of the constraint.

After having presented the quantitative results and the discussion with two versions of ChatGPT about the results, we now return to the research question of this paper.

6 Discussion

The main research question of this paper was: Can ChatGPT be used as a linguistic informant via its translations of literary texts? I have tried to answer this research question in two ways.

First, I have presented the results of the comparison of the HTs' and MT's translations of 108 bare plural subject nouns taken from two Dutch novels. The results have shown that there is a large correspondence. Like the HTs, the MT makes a distinction between unmodified existential nouns and modified existential nouns. Like the HTs, the MT also makes a distinction between unmodified existential nouns and unmodified generic nouns. Furthermore, like the HTs, the MT distinguishes between unmodified and modified generic nouns.

Second, discussions with ChatGPT-5 and ChatGPT-4o about the translations, showed that ChatGPT has some metalinguistic knowledge. On a syntactic analytical level ChatGPT performed more poorly.[11] The chatbot could not tell what the exceptions to Suñer's (1982) Naked Noun Constraint are. It had difficulty in seeing the role of modification in the choice of the article.

[11] One of the reviewers points out that a simple chat does not suffice to support this claim. A more complete test with benchmarking would be needed, namely a systematic testing with different prompts, also compared to replies that different syntacticians would give. After collecting the dataset, metrics should be used, such as accuracy and recall. This would, however, be a separate study, which has to be left for future research. Therefore, in this study, the claims made in the theoretical linguistic literature presented in Section 3 and not linguistic judgments provided by linguists are used as the benchmarking set. A comparison has been made with respect to the four categories distinguished in Table 2. The evaluation of the comparison has been made on the basis of features used in the literature, such as "non-specific", "generic", "modification" or their variants. Since the benchmarking set is based on general remarks made by linguists in the literature on four sets of data, a detailed comparison in terms of accuracy and recall is not possible. Therefore, only a summary of the results in more general terms has been given at the end of Section 5. The complete discussions on which the analysis is based can be consulted on OSF, see fn. 6.

The combination of these results shows that ChatGPT has reached observational adequacy. It is able to translate bare subject nouns like human translators. Furthermore, ChatGPT can act like a language teacher and can tell why it made different choices, using metalinguistic terminology, although these explanations are not always perfect. In this respect, ChatGPT may differ from non-expert native speakers, who cannot always tell why linguistic choices are made. However, on a descriptive level ChatGPT performs more poorly, as has also been mentioned by Mulders and Ruys (2024). ChatGPT does not see the structural role of modification, as in Suñer's (1992) account of the use of indefinites for modified existential nouns or as in Longobardi's (2001) account of modified generic subjects with a quantificational generic interpretation. Since ChatGPT has not reached descriptive adequacy, it has not reached explanatory adequacy either. It does not generalize over the role of modification for the four Romance languages.

The lack of descriptive and explanatory adequacy does not mean that ChatGPT could not improve. Training could possibly help ChatGPT to learn the impact of modification on the choice of the article in a certain Romance language and training could possibly also learn ChatGPT that modification has an impact in other Romance languages as well. The necessity of training, making use of few-shot prompts as was done in Mulder and Ruys' (2024) study, would show, however, that ChatGPT's knowledge is much different from the knowledge possessed by human beings, who do not need training to be able to judge the grammaticality of certain structures.

Although ChatGPT performs more poorly on linguistic analysis, it may help linguists, however, with respect to data validation and data augmentation, as shown in this paper. On the one hand, it was shown that the translations by the HTs were quantitatively and qualitatively more or less reproduced. On the other hand, it was shown thar discussions with ChatGPT on the choices that it makes may also help the researcher. In both cases, however, ChatGPT's outputs have to be checked. In the first case by checking the translations of the MT against those of the HTs and in the second case by checking ChatGPT's knowledge against linguists' knowledge.

In Section 4 it was observed that bare and *des*/*dei*-subjects mainly occur in narrative texts, but that even there they are rarely found. Linguistic analyses have mainly been based on isolated examples created by the linguists themselves. The corpus of 108 preverbal subject nouns in Dutch and their translations into four Romance languages may therefore be an interesting source for linguistic analysis. One of the most striking findings of the results presented in the previous section is that the two types of translators mainly used the definite article in their translation of unmodified existential subjects. This possibility is absent in Table 1 and has rarely been observed in the linguistic literature, but see Zamparelli (2002) for Italian.

A limitation of this study may be that it did not provide any training to ChatGPT. This was, however, not the goal of the study. In a follow-up study it could be investigated how much training ChatGPT would need to reach descriptive and

explanatory power in the domain of article choice with preverbal subject nouns. Furthermore, it has to be noted that the classification of certain results has to be viewed with caution. It was not always clear if a modified noun had to be classified as existential or as generic. Another limitation is that for the answers by ChatGPT in the dialogues no precise benchmarking against linguists' assessments took place.

It should be noted that it was not the goal of this paper to check the correctness of the translations by the HTs. On the contrary, the HTs' translations have been taken as the baseline and the MT's translations have been checked against them. It has been shown, however, that the HTs' translations may show variation, as it is the case for the MT's translations. Within a category, such as unmodified existential or modified generic nouns, in a second round the translator could make a different choice, which was already mentioned in Section 5 for ChatGPT-5 versus ChatGPT-4. For this reason, in this study the human and the machine translations have been compared per category, but not in a one by one comparison per noun.

7 Conclusion

Researchers have used various methods to gain insight into the acceptability of sentences, such as Grammaticality Judgment Tasks and corpus studies. There are linguistic phenomena that cannot be judged in isolation, but which need a context. This makes a Grammaticality Judgment Task less suitable for such phenomena. There are also phenomena that essentially occur in narrative texts, such as literary works, but generally in a very limited way. This limitation coupled with the fact that in each literary work the linguistic phenomenon under study occurs in a different context, makes a cross-linguistic comparison difficult to make. Therefore, in this study, a translation method was chosen. Based on two Dutch source texts in which 108 bare preverbal subject nouns were found, their translation by the professional translators and by ChatGPT was analyzed. The results were evaluated against linguistic analyses that have been provided in the literature.

The main goal of this paper was to investigate if ChatGPT may help linguistics to do research. It has been shown that ChatGPT may help to validate and augment data that have already been provided by others or that, as a chatbot, ChatGPT may help the linguist to gain more insight into the reason for a linguistic choice.

In answer to the AI-literature it has been shown that ChatGPT is a "wonder of a machine", with an increasing performance. This study is therefore not aimed at criticizing ChatGPT. This study has, however, shown that, although ChatGPT reaches observational adequacy and shows some metalinguistic knowledge, it does not have reached (yet) descriptive and explanatory adequacy.

In future research it could be investigated how much training would be necessary to go beyond observational adequacy and to improve the chatbot's metalinguistic analyses and if such a training would help.

Conflicts of Interest

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

Acknowledgments

This paper was presented at the Workshop "*Natural language and AI: New perspectives for linguistic studies"* of the Österreichische Linguistiktagung 2024 (ÖLT). I thank the audience for the fruitful discussion and the organizers for their work as editors of this Special Issue of *AI-Linguistica*.

Many thanks also go to the reviewers of this paper, for their critical remarks and valuable suggestions for improvement.

I also acknowledge the use of ChatGPT-4o for the translation into four Romance languages of the submitted paragraphs from the two Dutch literary works and the use of ChatGPT-4o and ChatGPT-5 for a discussion of the choices made in the translation of the paragraphs that were submitted for that purpose.

References

Primary sources

Pfeijffer, Ilja Leonard. 2013. *La Superba*. Amsterdam: De Arbeiderspers.

Pfeijffer, Ilja Leonard. 2018. *La Superba* (Claudia Cozzi, Trans.; Italian edn.). Rome: Nutrimenti. (Original work published in 2013)

Pfeijffer, Ilja Leonard. 2018. *Grand Hotel Europa*. Amsterdam: De Arbeiderspers.

Pfeijffer, Ilja Leonard. 2020. *Grand Hotel Europa*. (Claudia Cozzi, Trans.; Italian edn.). Roma: Nutrimenti. (Original work published in 2018)

Pfeijffer, Ilja Leonard. 2021. *Grand Hotel Europa*. (Françoise Antoine, Trans.; French edn.). (Original work published in 2018)

Pfeijffer, Ilja Leonard. 2021. *Grand Hotel Europa*. (Gonzalo Fernández Gómez, Trans.; Spanish edn.). Barcelona: Acantilado. (Original work published in 2018)

Pfeijffer, Ilja Leonard. 2021. *Grand Hotel Europa*. (Maria Leonor Raven, Trans.; Portuguese edn.). Porto: Livros do Brasil. (Original work published in 2018)

Pfeijffer, Ilja Leonard. 2022. *Grand Hotel Europa*. (Michele Hutchison, Trans.; English edn.). London: 4th Estate. (Original work published in 2018)

Secondary literature

Al Rousan, Rafat & Jaradat, Raghad & Malkawi, Mona. 2025. ChatGPT translation vs. human translation: an examination of a literary text. *Cogent Social Sciences* 11 (1). 1–21. https://doi.org/10.1080/23311886.2025.2472916

Bender, Emily M. & Gebru, Timnit & McMillan-Major, Angelina & Shmitchell, Shmargaret. 2021. *On the dangers of stochastic parrots: Can language models be too big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 610–623. https://doi.org/10.1145/3442188.3445922

Bosveld-de Smet, Leonie. 1998. On Mass and Plural Quantification. The case of French *des*/*du*-NPs. Ph.D. Dissertation. Groningen: University of Groningen.

Bouma, Gosse & van Noord, Gertjan & Malouf, Robert. 2001. Alpino: Wide-coverage computational analysis of Dutch. *Language and Computers* 37 (1). 45–59. https://doi.org/10.1163/9789004333901_004

Brown, Tom B. & Mann, Benjamin & Ryder, Nick & Subbiah, Melanie & Kaplan, Jared & Dhariwal, Prafulla & Neelakantan, Arvind & Shyam, Pranav & Sastry, Girish & Askell, Amanda et al. 2020. Language models are few-shot learners. In Larochelle, Hugo & Ranzato, Marc'Aurelio & Hadsell, Raia & Balcan, Maria-Florina & Lin, Hsuan-Tien (eds.), *Advances in Neural Information Processing Systems* 33. [Annual Conference on Neural Information Processing Systems 2020], 1877–1901. Red Hook, NY: Curran Associates Inc.

Cardinaletti, Anna & Giuliana Giusti. 2018. Indefinite determiners: Variation and optionality in Italo-Romance. In D'Alessandro, Roberta & Pescarini, Diego (eds.), *Advances in Italian Dialectology. Sketches of Italo-Romance Grammars*, 135–161. Leiden: Brill.

Carlson, Greg N. 1977. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1 (3). 413–456. https://doi.org/10.1007/BF00353456

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam & Roberts, Ian & Watumull, Jeffrey. 2023. The false promise of ChatGPT. *The New York Times*, 8 March 2023. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.

Cranenburgh, Andreas van. 2016. Rich Statistical Parsing and Literary Language. Ph.D. Dissertation. Amsterdam: University of Amsterdam.

Delfitto, Denis & Schroten, Jan. 1991. Bare plurals and the number affix in DP. *Probus* 3 (2). 155–185. https://doi.org/10.1515/prbs.1991.3.2.155

Dentella, Vittoria & Günther, Fritz & Murphy, Elliot & Marcus, Gary & Leivada, Evelina. 2024. Testing AI on language comprehension tasks reveals insensitivity to underlying meaning. *Scientific Reports* 14, 28083. https://doi.org/10.1038/s41598-024-79531-8

Englebert, Annick. 1992. *Le petit mot DE. Etude de sémantique historique*. Droz: Geneva.

Gibson, Edward & Evelina Fedorenko. 2013. *The need for quantitative methods in syntax and semantics research. Language and Cognitive Processes* 28 (1/2). 88–124. https://doi.org/10.1080/01690965.2010.515080

Hoop, de Helen. 1992. Case Configuration and Noun Phrase Interpretation. Ph.D. Dissertation Groningen: University of Groningen.

Katzir, Roni. 2023. Why large language models are poor theories of human linguistic cognition: a reply to Piantadosi. *Biolinguistics* 17. *Article e13153.* https://doi.org/10.5964/bioling.13153

Klis, Martijn van der & Le Bruyn, Bert & De Swart, Henriëtte. 2022. A multilingual corpus study of the competition between PAST and PERFECT in narrative discourse. *Journal of Linguistics* 58. 423–457. https://doi.org/10.1017/s0022226721000244

Lappin, Shalom. 2024. Assessing the strengths and weaknesses of Large Language Models. *Journal of Logic, Language and Information* 33. 9–20. https://doi.org/10.1007/s10849-023-09409-x

Leonetti, Manuel. 2013. Information structure and the distribution of Spanish bare plurals. In Kabatek, Johannes & Wall, Albert (eds.), *New Perspectives on Bare Noun Phrases in Romance and Beyond*, 121–155. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.141.05leo

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in Syntax and Logical Form. *Linguistic Inquiry* 25 (4). 609–665.

Longobardi, Giuseppe 2001. How comparative is semantics? A unified parametric theory of bare nouns and proper names. *Natural Language Semantics* 9 (4). 335–369. https://doi.org/10.1023/A:1014861111123

Meurers, W. Detmar. 2005. On the use of electronic corpora for theoretical linguistics. *Lingua* 115 (11). 1619–1639. https://doi.org/10.1016/j.lingua.2004.07.007

Moro, Andrea & Greco, Matteo & Stefano F. Cappa. 2023. Large languages, impossible languages and human brains. *Cortex* 167. 82–85. https://doi.org/10.1016/j.cortex.2023.07.003

Mulders, Iris & E.G. Ruys. 2024. ChatGPT as an informant. *Nota Bene* 12. 242–260. https://doi.org/10.1075/nb.00015.mul

Müller, Ana & Oliveira, Fátima. 2004. Bare nominals and number in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 3. 9–36. https://doi.org/10.5334/jpl.17

Myers, James. 2009. Syntactic judgment experiments. *Language and Linguistics Compass* 3 (1), 406–423. https://doi.org/10.1111/j.1749-818x.2008.00113.x

Oliveira Fátima & Silva Fátima. 2007. Indefinites and bare nouns in generic contexts in European Portuguese. *Verbum* 29 (3/4). 225–241.

Oosterhof, Albert. 2008. *The Semantics of Generics in Dutch and Related Languages*. Amsterdam: John Benjamins. https://doi.org/10.1075/la.122

OpenAI. 2021. ChatGPT (Versions GPT-4o and GPT-5) [Computer software]. https://openai.com/

Piantadosi, Steven. 2024. Modern language models refute Chomsky's approach to language. In Gibson, Edward & Poliak, Moshe (eds.), *From Fieldwork to Linguistic Theory: A tribute to Dan Everett*, 353–414. Berlin: Language Science Press.

Roodenburg, Jasper. 2004. Pour une approche scalaire de la déficience nominale: La position du français dans une théorie des <<noms nus>>". Utrecht: LOT.

Rossi, Sarah & Formichi, Guido & Neri, Sofia & Sgrizzi, Tommaso & Zanollo, Asya & Bressan, Veronica & Chesi, Cristiano. 2025. Acquisition in babies and machines: Comparing the learning trajectories of LMs in terms of syntactic structures (ATTracTSS Test Set). In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, 1007–1017.

Salem, Murad. 2010. Bare nominals, information structure and word order. *Lingua* 120. 1476–1501. https://doi.org/10.1016/j.lingua.2009.11.002

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press

Schütze, Carson T. 2011. Linguistic evidence and grammatical theory. *WIREs Cognitive Science* 2. 206–221. https://doi.org/10.1002/wcs.102

Schütze, Carson T. 2016. *The Empirical Base of Linguistics: Grammaticality judgments and linguistic methodology*. [Classics in Linguistics 2]. Berlin: Language Science Press. https://doi.org/10.17169/langsci.b89.100

Schütze, Carson T. & Sprouse, Jon. 2014. Judgment data. In Podesva, Robert J. & Sharma, Devyani (eds.), *Research Methods in Linguistics*, 27–50. Cambridge: Cambridge University Press.

Soares, Nuno Verdial. 2018. *Bare Nouns in European Portuguese*. Dissertation Lisbon: University of Lisbon.

Stefanowitsch, Anatol. 2011. Cognitive linguistics meets the corpus. In Brdar, Mario, Gries, Stefan Th. & Fuchs, *Fuchs, Žic* (eds.), *Cognitive Linguistics. Convergence and Expansion*. [*Human Cognitive Processing* 32], 257–289. Amsterdam: John Benjamins. https://doi.org/10.1075/hcp.32.16ste

Suñer, Margarita. 1982. *Syntax and Semantics of Spanish Presentational Sentence-Types*. Washington D.C.: Georgetown University Press.

Wall, Albert. 2017. *Bare Nominals in Brazilian Portuguese – An Integral Approach*. Amsterdam: John Benjamins.

Warstadt, Alex & Parrish, Alicia & Liu, Haokun & Mohananey, Anhad & Peng, Wei & Wang, Sheng-Fu & Bowman, Samuel R. 2020. BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics* 8. 377–392. https://doi.org/10.1162/tacl_a_00321

Warstadt, Alex & Mueller, Aaron & Choshen, Leshem & Wilcox, Ethan & Zhuang, Chengxu & Ciro, Juan & Mosquera, Rafael & Paranjabe, Bhargavi & Williams,

Adina & Linzen, Tal & Cotterell, Ryan. 2023. Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora. In Warstadt, Alex & Mueller, Aaron & Choshen, Leshem & Wilcox, Ethan & Zhuang, Chengxu & Ciro, Juan & Mosquera, Rafael & Paranjabe, Bhargavi & Williams, Adina & Linzen, Tal & Cotterell, Ryan (eds.), *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, 1–34, Singapore: Association for Computational Linguistics. https://doi.10.18653/v1/2023.conll-babylm.1

Zamparelli, Roberto. 2002. Definite and bare kind-denoting noun phrases. In Beyssade, Claire & Bok-Bennema, Reineke & Drijkoningen, Frank & Monachesi, Paola (eds.), *Romance Languages and Linguistic Theory 2000*, 305–343. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.232.17zam