Khazar Journal of Humanities and Social Sciences

Volume 28 | Issue 4 Article 5

12-31-2025

# Lost in Ambiguity: AI and the Limits of Processing Turkish Morphology

Engin Evrim Önem
*Erciyes University, School of Foreign Languages, Kayseri, Turkiye*, eonem@erciyes.edu.tr



## How to Cite This Article

Önem, Engin Evrim (2025) "Lost in Ambiguity: AI and the Limits of Processing Turkish Morphology," *Khazar Journal of Humanities and Social Sciences*: Vol. 28: Iss. 4, Article 5.
Available at: https://kjhss.khazar.org/journal/vol28/iss4/5

KHAZAR JOURNAL OF HUMANITIES AND SOCIAL SCIENCES 2025;28:294–308

**ORIGINAL STUDY**

# Lost in Ambiguity: AI and the Limits of Processing Turkish Morphology

**Engin Evrim Önem**

Erciyes University, School of Foreign Languages, Kayseri, Turkiye

## ABSTRACT

Despite rapid progress in natural language processing, artificial intelligence (AI) systems continue to struggle with languages characterized by high morphological complexity, such as Turkish. A central difficulty lies in their limited ability to resolve ambiguity, which is deeply embedded in Turkish through features like flexible word order, extensive case marking, and agglutinative structures. Current AI models, primarily shaped by exposure to massive corpora of performance data, excel at detecting surface-level statistical regularities but fall short of grasping the grammatical principles and semantic nuances that underpin genuine linguistic competence. This reliance on performance over competence makes disambiguation especially problematic, as the systems lack the deep structural awareness required to interpret sentences where multiple readings are possible. To explore these weaknesses, the study tested five prominent AI models, Gemini, Claude, ChatGPT, Deepseek, and Grok, using ten carefully selected ambiguous Turkish sentences. Without being alerted to the ambiguity, the systems frequently generated interpretations that were semantically inappropriate, inconsistent, or distorted by hallucinated content. The findings illustrate how data-driven training alone cannot equip AI with the pragmatic reasoning and world knowledge necessary for accurate interpretation. The paper argues for a shift toward models that integrate richer linguistic theory, enabling AI to move beyond statistical mimicry toward a more human-like capacity for language understanding. Such an approach is vital for developing tools that can handle morphologically rich and ambiguity-prone languages with greater fidelity.

**Keywords:** Artificial intelligence (AI), Large language models (LLMs), Disambiguation, Turkish, Linguistic competence, Linguistic performance, Limitations of AI

## Introduction

Artificial intelligence (AI) is a field of computer science that aims to create computer systems that can perform tasks that typically require human intelligence. These tasks include speech recognition, natural language processing (NLP), text generation and translation, video, sound, and image generation, as well as decision-making and planning.

Received 25 August 2025; revised 19 December 2025; accepted 9 January 2026.
Available online 31 December 2025

E-mail address: eonem@erciyes.edu.tr (E. E. Önem).

*https://doi.org/10.5782/2223-2621.1583*

AI systems imitate human behavior and decision-making processes by processing large datasets and identifying patterns to make decisions or take actions. While modern AI has made considerable strides in interpreting and generating human language, these advancements bring forth significant implications for linguistics and communication, particularly concerning their ability to handle the intricate nuances of human language.

The confluence of AI and linguistics has been particularly fertile, with each field informing and enriching the other. As linguistic models become increasingly sophisticated, they analyze linguistic data and reveal something about the structure and evolution of human language. Interestingly, although AI has been widely used in linguistic studies, linguists still lack a consensus on the role of AI in linguistic research. For example, Alaqlobi *et al.* (2024) analyze 73 articles from a linguistic research perspective and report that AI models, such as ChatGPT, are considered promising for linguistic tasks. However, responsible implementation is crucial for ensuring ethical and quality assurance. Apart from the ethical issues surrounding AI, the distinction between human intelligence and artificial intelligence in terms of linguistic competence and performance is also considered a barrier (Dupre, 2021; McShane & Nirenburg, 2021).

Apart from this barrier, the current artificial intelligence can be referred to as "weak AI," which is designed to simulate human intellectual processes, while "strong AI" is designed to develop genuine cognition, a form of intelligence that indeed has and knows the mental states (Flowers, 2019; Hockly, 2023). Consequently, whereas current AI inherently possesses some limitations, the eventual deployment of strong AI, as acknowledged by Dupre (2021), could enable AI systems to perform linguistic functions on par with the human species, along with clear insight into operationalization that could create new research paradigms in defining human cognitive processes. Indeed, Sindhu *et al.* (2024) present potential future trends in AI technologies, including enhanced context handling capabilities, methods for bias mitigation, and increased computational efficiency through emerging trends such as federated learning. Therefore, numerous unanticipated AI improvements are looming on the horizon.

In the study of language processing, ambiguity is a crucial concept, and it is defined as a property of a word, phrase, or sentence that is open to multiple interpretations. The immediate challenge for both human and artificial language processors is disambiguation, the cognitive or computational process of resolving this uncertainty by selecting the single, contextually appropriate meaning. Evaluating an AI model's effectiveness in disambiguating complex Turkish sentences allows us to assess its linguistic competence, which is the theoretical, underlying knowledge of grammatical rules and structures, against its linguistic performance, which is the actual use of language in concrete situations. This paper aims to demonstrate how Artificial Intelligence (AI) is challenged by modern linguistic research, particularly when confronting the inherent ambiguities of morphologically rich languages like Turkish, and why it is essential to consider its linguistic limitations, especially in relation to the distinction between linguistic performance and competence. The outline of the paper is as follows: after clarifying often-confused AI concepts, the inherent linguistic challenges that hinder AI's full potential in disambiguation, particularly in Turkish, are explored. The methodology involves presenting ambiguous Turkish sentences to LLMs without informing them of their ambiguous nature, analyzing how these systems interpret complex linguistic structures through implicit processing. Through examples from this analysis, the paper examines the underlying reasons for disambiguation failures, exploring how the agglutinative morphology, flexible word order, and rich semantic structures of Turkish pose challenges for current AI capabilities. The analysis reveals systematic patterns in how LLMs handle, or mishandle, ambiguous input across different linguistic contexts. Finally, the conclusion synthesizes these findings to highlight the broader implications of AI language processing in morphologically complex languages, such as Turkish.

## Defining AI: From technical distinctions to contemporary usage

While the technical distinctions between AI, ML, generative AI, and LLMs are essential to understand, in contemporary discourse and practical applications, the term "artificial intelligence" has become predominantly associated with Large Language Models. When people refer to "AI" in everyday contexts, whether discussing ChatGPT, Claude, or similar systems, they are typically referring to LLMs specifically. This linguistic shift reflects the prominent role that language models have assumed in the public perception of artificial intelligence. Therefore, throughout this paper, "AI" will be used as an umbrella term that primarily encompasses LLMs, acknowledging that while AI is technically a broader concept encompassing various machine learning approaches, current usage patterns have made it synonymous with these language-focused generative systems. This terminological choice aligns with how the field is commonly understood and discussed in both academic and popular contexts, where LLMs have become the most visible and widely deployed manifestation of artificial intelligence technology.

Artificial intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence, including learning from experience, recognizing patterns, understanding language, solving problems, making decisions, and adapting to new situations (Prodan *et al.*, 2024). These systems perform such tasks via machine learning: deep learning, computer vision, and reinforcement learning. Machine learning (ML) is a subset of AI where systems learn and improve from experience without explicit programming via algorithms to train AI models on vast datasets to identify patterns and make predictions. In other words, data exposure improves systems using neural networks with multiple layers. Machine learning employs commercial-grade algorithms that enable AI models to extract functional patterns from large datasets. Such algorithms are learning mechanisms that enable the system to enhance its performance over time on specific tasks through repetitive exposure to examples. For instance, while training an image recognition system, the model is exposed to thousands of labeled images, learning to extract visual features that characterize various objects over time. Neural networks, particularly deep networks with numerous layers, are the architectural backbone of most modern ML systems. The "deep" in deep learning is the chain of hidden layers that enables the system to build increasingly abstract data representations. Each later layer relies on the earlier one, and the network can pick up hierarchical patterns - from low-level edges and shapes to more abstract objects and concepts. The learning process is adapting the weights connecting neurons according to feedback received about prediction performance. The optimization process, typically guided by methods such as gradient descent, progressively reduces the discrepancy between predicted and actual outcomes. The most potent aspect of machine learning is its ability to generalize from trained data to unseen situations. Trained adequately, an ML system can provide logical predictions on new data, enabling applications in computer vision and natural language processing. In addition, the more diverse and extensive datasets systems are exposed to, the higher predictive accuracy they are likely to exhibit, leading to improved performance in various fields of application. Consequently, while ML is a method for training those machines, AI is the broader concept of intelligent machines.

In terms of output, a distinction is made between discriminative and generative types of AI. Through learning the existing patterns, generative AI models can synthesize new data that mimics the source material (Hagos *et al.*, 2024). In other words, generative AI is a highly sophisticated type of machine learning that extends beyond recognizing patterns to generating patterns. These models learn to identify patterns in data and generate entirely new material that maintains the statistical and structural characteristics of the training data. On the other hand, discriminative AI is the opposite of generative AI in machine learning since discriminative AI analyzes and/or acts upon existing data, focusing on boundary determination and classification rather than creation (Corchado *et al.*, 2023). These systems are adept at analyzing existing data to predict, classify, or determine something based on patterns within the data. It can be said that discriminative AI learns boundaries between data categories for classification tasks. At the same time, generative AI creates new data similar to its training data, which results in generating images or videos (e.g. DALL-E), computer codes (e.g. Anthropic Claude), music (e.g. beatoven.ai), or texts (e.g. ChatGPT).

Large Language Models (LLMs) are a specific type of generative AI built upon foundation models and focused on natural language processing. In other words, LLMs are a particular application within generative AI, explicitly focusing on language-based tasks (Corchado *et al.*, 2023). Initially, Large Language Models (LLMs) are based on Language Models (LMs) that aim to predict the next word or character in a given sequence of text, thereby developing algorithms and models that can understand and generate coherent human language (Hadi *et al.*, 2023). However, as noted by Hadi *et al.* (2023), the breakthrough for LLMs occurred in 2017 with the advent of a new architecture that facilitated parallelization and improved the handling of long-distance relationships within text, as well as the effect of context. Thereafter, the architectural design continued to evolve, but one underlying element of LLMs, such as the GPT and BERT families, remained the same: their training process. It consists of initial pre-training on general text data and subsequent fine-tuning for targeted tasks (Sindhu *et al.*, 2024). LLMs learn linguistic structures by being trained on massive data, enabling them to produce fluent, contextually relevant text and understand the relationship between words. Later, the models are fine-tuned, whereby they are trained on specialized data to enable them to perform exceptionally well in conversation and text generation. As a result, LLMs have demonstrated exceptional performance in a wide range of tasks across various fields. For example, Sindhu *et al.* (2024) believe that generative AI models can be applied in various fields, including medicine, education, science, mathematics, finance, and robotics. Similarly, Hagos *et al.* (2024) also enumerate the prospective applications of generative AI in drug and material discovery, fraud detection, and content personalization. In terms of the language-specific applications of generative AI, in addition to text generation, translation, summarization, question-answering, and sentiment analysis (McShane & Nirenburg, 2021), language understanding, speech recognition, and chatbots (Hagos *et al.*, 2024) are the most prevalent AI applications in language.

## Methodology

To evaluate how Large Language Models (LLMs) perform in Turkish sentence disambiguation, a corpus of ambiguous sentences to represent the primary types of ambiguity found in Turkish, such as Lexical Ambiguity, Morphological Ambiguity, and various forms of Structural/Syntactic Ambiguity (including its sub-types like scope and relative clause attachment), was created. Then, the sentences were reviewed and confirmed by a linguist in terms of the varieties of ambiguity they reflect, as well as in their ability to isolate and exemplify distinct forms of linguistic ambiguity, making them reliable test cases for evaluating an AI model's ability to recognize and process these challenges in Turkish. The list of sentences and the type of ambiguity they reflect is given in Table 1 below.

The rationale for the sentence selection was to ensure comprehensive coverage of the major ambiguity types in Turkish while strategically maximizing the testing potential of Structural Ambiguity. First, for the lexical ambiguity, three sentences were chosen which primarily probe the AI's semantic knowledge and reliance on external context to select the correct homonymous meaning (e.g., superhero vs. weight unit; 'wash' vs. 'demolish') to test the AI's ability to integrate real-world knowledge with word-level interpretation. Regarding the morphological ambiguity, two sentences were created to test the AI's ability to process Turkish's complex agglutinative suffixes. Resolving these requires distinguishing between identical morphemes that signal different grammatical roles, which is a challenge specific to processing high-agglutinative languages. Finally, since structural ambiguity is inherently complex, as it involves multiple possible *syntactic tree structures* derived from the same sequence of words, making it a powerful tool for diagnosing whether an AI operates based on human-like constraint-based processing (which prioritizes context) or purely on simplistic structural heuristics (like Late Closure or Recency), five sentences were created for structural ambiguity. Those sentences included Relative Clause Attachment, Quantifier Scope, and Comparison/Negation Scope, which are where the parsing strategy of the human brain most explicitly contrasts with purely structural computational models. Testing different variations of scope and attachment provides the richest dataset for comparing the AI's "default parse" against the established human baseline preference (e.g., the NP1 preference in neutral RC contexts).

In the end, the corpus included ten sentences that are not merely illustrations but rather diagnostic probes designed to systematically measure how AI handles phenomena known to challenge human language processing in Turkish, by limiting the corpus to ten sentences, a highly consistent and controlled battery of tests allowed for the deliberate inclusion of examples representing the three major, theory-driven ambiguity types

**Table 1.** Sentences and type of ambiguity.

| Ambiguity Type | Sentences |
|---|---|
| Lexical | insan-lar Türkiye-de Batman-ı sev-er-ler]<br>person-PL Turkey-LOC Batman-ACC like-AOR-3PL<br>onlar bura-yı yı k-ar-lar<br>they here-ACC demolish-AOR-3PL<br>suçlu-lar-ı n yüz-ü ası l-dı<br>criminal-PL-GEN face-3SG.POSS sullen-PST.3SG |
| Morphological | gel-di-ğin-i duy-ma-dı m<br>come-PST-NMLZ.3SG.POSS-ACC hear-NEG-PST.1SG<br>yaşlı adam-ı döv-dü<br>old man-ACC beat-PST.3SG<br>adam ol baban gibi eşek ol-ma<br>man be father-POSS.2SG like donkey be-NEG |
| Structural | Ali benim gibi çalışkan değil<br>Ali 1SG-GEN like hardworking NEG.COP<br>ihtiyar hasta eş-i-nin yan-ı -ndan<br>elderly sick spouse-3SG.POSS-GEN<br>side-3SG.POSS-ABL<br>ayrı l-ma-dı<br>leave-NEG-PST.3SG<br>adam önce karı -sı -nı sonra kı z-ı -nı<br>man first wife-3SG.POSS-ACC then<br>daughter-3SG.POSS-ACC<br>gör-dü<br>see-PST.3SG<br>herkes-in çocuğ-u-nun başarı -sı -nı<br>everyone-GEN child-3SG.POSS-GEN<br>success-3SG.POSS-ACC<br>gör-me-si-ni iste-r-im<br>see-NMLZ-3SG.POSS-ACC want-AOR-1SG |

relevant to Turkish psycholinguistics. Specifically, this number was deemed sufficient to allocate sentences to test distinct structural subtypes, such as Relative Clause Attachment and Quantifier/Comparison Scope, which are crucial for probing both human and AI parsing strategies. By systematically including examples that require the integration of lexical, morphological, and structural information, the study effectively measures the AI's competence against human processing strategies, particularly its crucial ability to use context to resolve complex structural indeterminacies.

It's essential to understand that while many of the ambiguous Turkish sentences in this study exhibit overlapping types of ambiguity (e.g., a sentence may contain both a lexical ambiguity and a structural one), the author has chosen to group and treat them based on the primary linguistic mechanism being investigated or the most salient ambiguity type used for testing the AI models. For instance, sentences like "yaşlı adamı dövdü" are fundamentally ambiguous due to both Morphological (subject pro-drop/agreement) and Structural (referential) factors. However, the author chose to primarily group the comparative sentence "adam ol baban gibi eşek olma" under Structural Ambiguity because the core research question being explored is the AI's ability to handle Scope Ambiguity (the unclear domain of the comparison "baban gibi", rather than the word-level idiomatic ambiguity that is also present. Similarly, although "suçluların yüzü asıldı" has a lexical ambiguity in "yüz", it's also morphologically complex. By prioritizing the most theoretically interesting or experimentally challenging ambiguities, such as the various forms of Structural Ambiguity that directly test parsing models, the author ensures a clear and diagnostic framework for comparing AI performance against established human psycholinguistic findings. The grouping reflects the experimental focus, not an absolute or exhaustive linguistic classification.

To evaluate how LLMs perform in Turkish sentence disambiguation, ten ambiguous sentences were presented to AI systems across five different models randomly (Gemini, Claude, ChatGPT, Deepseek, and Grok - two instances each). The language models were prompted to interpret ambiguous sentences in Turkish without being notified of their ambiguous nature (implicit processing). A prompt preceded each sentence, "Explain the meaning of this sentence: [sentence]", and the responses generated by different LLMs were analyzed to understand how these systems handle complex linguistic structures when unaware of potential ambiguities.

## *AI responses to Turkish ambiguous sentences*

The following section presents responses generated by AI systems when processing Turkish ambiguous sentences directly. The analysis reveals systematic patterns in how these systems interpret, or misinterpret, complex linguistic structures.

### *Sentence 1: "yaşlı adamı dövdü"*

***Response generated by Gemini:***. "The Turkish sentence 'yaşlı adamı dövdü' translates to 'He/She/It beat the old man' or 'The old man was beaten (by him/her/it).'

Here's a breakdown of the words:

- yaşlı : old (adjective)
- adamı : the man (noun in the definite accusative case, meaning 'the man' is the direct object of the action)
- dövdü: beat/hit (past tense verb, third person singular - 'he/she/it beat')

The sentence implies that someone (a third person singular subject, which is often omitted in Turkish if clear from context) acted beating an old man."

*Sentence 2: “adam ol baban gibi eşek olma”*

***Response generated by Gemini:***. “The Turkish sentence ’adam ol baban gibi eşek olma’ is a rather harsh and culturally specific admonishment that translates literally to: ’Be a man like your father; don’t be a donkey.’

However, to understand its true meaning, we need to look beyond the literal translation:

- ’Adam ol’ (Be a man): This isn’t just about gender. In Turkish culture, ’adam olmak’ implies being responsible, honorable, mature, strong, and respectable.
- ’Baban gibi’ (Like your father): This part suggests that the father is being held up as an example of these desired qualities.
- ’Eşek olma’ (Don’t be a donkey): In this context, ’eşek’ (donkey) is used as a strong insult, implying someone is stupid, foolish, stubborn, or obstinate.

The sentence is a severe scolding, typically directed at someone who is acting foolishly or irresponsibly. The speaker is essentially saying: ’Act responsibly and wisely, like your father does. Don’t be foolish or ignorant.’”

*Sentence 3: “Ali benim gibi çalışkan değil”*

***Response generated by Claude:***. “This Turkish sentence means ’Ali is not as hardworking as me’ or ’Ali is not hardworking like me.’

Breaking it down:

- Ali = Ali (a person’s name)
- benim gibi = like me / as me
- çalışkan = hardworking, diligent
- değil = is not

The sentence structure uses a comparison format where someone is comparing Ali’s work ethic to their own, stating that Ali doesn’t work as hard as they do. The phrase ’benim gibi’ (like me) is the key comparative element here.”

*Sentence 4: “geldiğini duymadım”*

***Response generated by Claude:***. “This Turkish sentence means ’I didn’t hear that you came’ or ’I didn’t hear that he/she came.’

Breaking it down:

- geldiğini = ’that (you/he/she) came’ - this is the accusative form of the verbal noun ’geldiği’ (coming)
- duymadı m = ’I didn’t hear’ - the negative past tense of ’duymak’ (to hear)

The sentence structure follows Turkish grammar where the object clause comes before the main verb. The ambiguity about whether it refers to ’you’ or ’he/she’ is normal in Turkish, as the context would usually make this clear in conversation.”

*Sentence 5: “herkesin çocuğunun başarısını görmesini isterim”*

***Response generated by ChatGPT:***. “The sentence ’Herkesin çocuğunun başarısını görmesini isterim’ is in Turkish and can be translated as: ’I wish that everyone sees their child’s success.’

Meaning Breakdown:

- Herkesin: everyone’s
- çocuğunun: their child’s
- başarısını : success (of the child)

- görmesini: the act of seeing
- isterim: I want / I wish

Interpretation: The speaker is expressing a benevolent wish: that every person will live to witness and experience the success of their own child. It reflects a positive, empathetic sentiment often found in well-wishing or blessings."

*Sentence 6: "adam önce karısını sonra kızını gördü"*

***Response generated by ChatGPT:***. "The Turkish sentence 'Adam önce karısını sonra kızını gördü' translates to: 'The man first saw his wife, then his daughter.'

Meaning Breakdown:

- Adam: the man
- önce: first
- karısını : his wife
- sonra: then
- kızını : his daughter
- gördü: (he) saw

Interpretation: On the surface, the sentence means that a man saw two people in sequence: first his wife, then his daughter."

*Sentence 7: "ihtiyar hasta eşinin yanından ayrılmadı"*

***Response generated by Deepseek:***. "The sentence 'İhtiyar hasta eşinin yanından ayrılmadı' in Turkish translates to English as: 'The elderly man did not leave his sick wife's side.'

Breakdown of the sentence:

- İhtiyar – elderly, old (referring to a man in this context)
- hasta eşinin – his sick spouse (wife/husband, but 'eş' often implies 'wife' unless specified otherwise)
- yanı ndan – from (her/his) side
- ayrı lmadı – did not leave

So, it describes an elderly man who stayed by his sick wife's side without leaving. It conveys care, devotion, and possibly a sense of duty or love."

*Sentence 8: "onlar burayı yıkarlar"*

***Response generated by Deepseek:***. "The sentence 'Onlar burayı yıkarlar.' is in Turkish and translates to 'They destroy this place.' or 'They will destroy this place.' in English.

Breakdown:

- Onlar → 'They' (3rd person plural)
- burayı → 'this place' (accusative case, indicating the direct object)
- yıkarlar → 'destroy' (present simple tense, 3rd plural)

Interpretation:

- Present Habitual Meaning: 'They destroy this place.' (repeated action, e.g., invaders, vandals)
- Future Meaning (depending on context): 'They will destroy this place.' (a prediction or warning)."

*Sentence 9: "insanlar Türkiyede Batman'ı severler"*

***Response generated by Grok:***. "The sentence 'İnsanlar Türkiye'de Batman'ı severler' is in Turkish and translates to 'People in Turkey love Batman' in English.

Here's a breakdown of its meaning:

- İnsanlar: Means 'people.'
- Türkiye'de: Means 'in Turkey.'
- Batman'ı : Refers to 'Batman,' likely the superhero character, as the suffix '-ı' indicates the accusative case.
- Severler: Means 'they love,' the third-person plural form of the verb 'sevmek' (to love)

The sentence expresses that Batman, presumably the comic book character or related media, is popular or well-liked among people in Turkey."

*Sentence 10: "suçluların yüzü asıldı"*

***Response generated by Grok:***. "The sentence 'suçluların yüzü asıldı' is in Turkish and translates to 'the faces of the guilty were hung' or more idiomatically, 'the guilty looked dejected' in English.

Here's a breakdown of its meaning:

- suçluların: This is the genitive plural form of 'suçlu,' meaning 'the guilty' or 'the culprits.'
- yüzü: Means 'face' or 'expression'
- asıldı : The past tense of the verb 'asılmak,' which in this context means 'to be hung' or 'to droop.' When used with 'yüz' (face), it's an idiomatic expression meaning someone looks sad, dejected, or ashamed.

The sentence conveys that the guilty people displayed expressions of sadness, shame, or disappointment, likely because they were caught, confronted, or punished."

### *Where AI fails: Disambiguation challenges in Turkish*

In the general sense, apart from more technical problems, including domain adaptation, computational cost, limited context window, long-term memory, measuring capability, and quality (see Hagos *et al.*, 2024, for a comprehensive list of concerns), the fundamental difference between linguistic *performance* and *competence* stands out. Linguistic competence is the internalized, unspoken knowledge the speaker has of language structure and rules - what they know automatically about how their language works.

In contrast, linguistic performance refers to the actual use and comprehension of language in everyday life, which can be influenced by various factors such as attention, memory constraints, distractions, and errors. As Dupre (2021) states, while language models and GenAI are trained on data based on linguistic performance, theoretical theories focus on competence-based data. Dupre (2021) continues to explain how performance reflects competence, which sets "a barrier" for AI, and concludes that the nature of linguistic competence is beyond the scope of insights provided by AI models. While all advancements enable AI, particularly LLMs, to generate fluent and seemingly coherent text, their underlying mechanism relies heavily on identifying statistical patterns from vast datasets of linguistic *performance*. This reliance on performance data, rather than an inherent understanding of linguistic *competence*, forms the crux of the challenges AI faces in handling complex linguistic phenomena such as disambiguation, especially in morphologically rich languages like Turkish.

Turkish, as an agglutinative language with flexible word order and extensive case marking, presents unique and significant challenges for AI systems trained primarily on performance data. The morphological richness often leads to multiple interpretations of a single word or sentence, where human speakers effortlessly disambiguate based on context and internalized grammatical knowledge. Most of the challenges AI faces when disambiguating Turkish sentences are due to the complex morphological and syntactic properties of Turkish.

### *Morphological complexity and agglutination*

Turkish can pack enormous amounts of grammatical information into single words through suffixation. For example, in "yaşlı adamı dövdü", the morphology doesn't distinguish between subject and object when both are animate and human. AI systems fail because they cannot determine whether "the old man beat [someone]" or "[someone] beat the old man" without pragmatic reasoning about plausibility, which is related to the agglutinative nature of the Turkish language. However, another interesting point stands out when considering AI in "adam-ı" as a possessive suffix and relating it to *his man/her man*, making "yaşlı adamı" equivalent to "the old man's [something/someone]." This interpretation, while grammatically possible in extended contexts, demonstrates how AI systems can generate overly complex analyses that native speakers would not typically consider.

Similarly, a clause like "gel-di-ğin-i" contains multiple morphemes (-DI past tense, -GIN-relativizer, -(y)I accusative case marker), and AI systems often struggle to correctly parse these layered meanings, especially when multiple interpretations are morphologically valid. The assumptions the AI system made to disambiguate the clause were partly correct. However, Turkish does not make a distinction between the third person singular in terms of animacy. In contrast, the AI system focused on that aspect and missed the morphological similarity between "his/her/its coming" and "your coming," which represents a fundamental misunderstanding of Turkish morphology.

### *Homophonous suffixes and semantic selectional restrictions*

Many Turkish suffixes are homophonous but serve different functions, and "suçluların yüzü asıldı" demonstrates this perfectly as "yüz" means both "face" and "hundred." In contrast "asıldı" means "was hanged" or "disappointed" when used with "yüz." This created various possible interpretations, but AI lacked the semantic selectional restrictions to determine that faces don't get executed while numbers don't have physical properties, and missed the secondary meaning concerning metonymy. Hence, a human effortlessly disambiguates the subject based on the preceding discourse or the real-world context of the interaction. AI, lacking this deep understanding of implicit referents and world knowledge, often defaults to a statistically probable translation or provides a generic one, failing to capture the specific intended meaning.

### *Cultural and pragmatic knowledge deficits*

Pragmatic as well as contextual knowledge is also key for interpreting ambiguous sentences. As for "adam ol baban gibi eşek olma", it required understanding Turkish cultural metaphors and pragmatic force. Although the disambiguation provided by the AI system appears valid, there is no doubt that AI systems often miss cultural mappings, potentially interpreting ambiguous sentences in a manner that differs significantly from how a native speaker of that language would do.

Also, “insanlar Türkiyede Batman’ı severler” causes significant AI failure because “Batman” could refer to the Turkish city or the superhero. Without cultural knowledge and proper noun disambiguation capabilities, AI systems often default to internationally known referents rather than considering geographical context. However, there was a phonological clue in that sentence concerning Turkish phonotactics. If “Batman” is pronounced as /batman/ (the city), due to the phonotactical constraints, the accusative case marker should be *-I*, but when it is pronounced as /bætmæn/ (the superhero), the accusative case marker should be *-İ*. The AI system failed to recognize this crucial phonological distinction.

### *Data representation and training limitations*

However, an AI system relying purely on statistical co-occurrences may struggle to differentiate without explicit contextual cues that precisely mirror its training and representation data, which is another reason why AI fails to disambiguate Turkish sentences. Turkish is relatively underrepresented in training corpora compared to English, meaning AI models encounter fewer examples of ambiguous constructions and their resolutions. This is particularly problematic with culturally specific expressions and regional proper nouns, such as “Batman.” Also, Turkish data may be skewed toward formal written language, while colloquial spoken Turkish (where much disambiguation occurs through prosody and gesture) remains underrepresented in training data.

In addition to that, Turkish’s rich morphological system paradoxically creates more ambiguity points requiring human-like reasoning. The language’s ability to encode precise grammatical relationships through morphology often results in multiple valid structural interpretations that can only be resolved through cultural knowledge, pragmatic inference, and sophisticated discourse processing capabilities that current AI architectures fundamentally lack.

### *Syntactic flexibility and scope ambiguities*

Syntactic flexibility, scope ambiguities, and morphological parsing complexity are other reasons why AI failed at disambiguating some sentences. Turkish’s relatively free word order creates parsing challenges. For instance, “adam önce karısını sonra kızını gördü” demonstrates temporal scope ambiguity: did he see his wife first, then his daughter, or did he see his wife’s daughter, then someone else’s? AI struggles with these nested temporal-possessive relationships because it cannot weigh the relative likelihood of different family relationship interpretations.

Another reason why AI failed was related to quantifier scope handling in “herkesin çocuğunun başarısını görmesini isterim”, which showed how multiple possessive chains created scope ambiguity. That sentence could mean “I want everyone’s child’s success” (broad scope) or “I want the success of each person’s child” (narrow scope). These are logically different but morphologically identical, requiring sophisticated quantifier-possessive interaction processing that AI systems lack.

Related to the morphological parsing complexity of Turkish, adjacency or scope was seen again in parsing “ihtiyar hasta eşinin yanından ayrılmadı”, which showed how three consecutive adjectives/nouns create parsing ambiguity. The morphological marking is insufficient, requiring knowledge of typical caregiving relationships in the world.

Finally, “Ali benim gibi çalışkan değil” illustrates comparison structure ambiguity, which could be interpreted as either “Ali is not as hardworking as I am” (scalar comparison) or “Ali is not as hardworking as I am” (manner comparison). AI failed to distinguish between these because it lacked pragmatic reasoning about typical communicative intentions.

### *Lexical ambiguity and tense-aspect interactions*

AI responded to “onlar burayı yıkarlar” by focusing on tense and aspect ambiguities as “yıkarlar” could be present or future tense, such as “they demolish” or “they will demolish this place.” However, the AI system failed to realize that “yıkmak” has multiple meanings, including both “wash” and “demolish” demonstrating another layer of lexical ambiguity that compounds the morphological complexity.

The core issue is that Turkish disambiguation requires integrating morphological analysis, syntactic parsing, pragmatic reasoning, cultural knowledge, and often prosodic information simultaneously. Current AI systems fail because they cannot perform this multimodal, context-sensitive processing at the level required for accurate disambiguation of morphologically rich, culturally embedded language.

### *Systemic patterns in AI failures*

The documented deficiencies in computational language processing systems reflect broader theoretical debates that permeate contemporary research. Research by Curry *et al.* (2024) demonstrates that automated semantic classification often produces overly broad categorizations that are unsuitable for domain-specific applications. Their findings reveal significant issues with erroneous conclusions and dataset manipulation, casting doubt on the fundamental trustworthiness of these technologies. In parallel work, Backus *et al.* (2023) highlight how neural language architectures generate fictitious content, termed “synthetic fabrications,” despite producing syntactically valid constructions. This phenomenon emerges because these systems extract statistical regularities from lexical co-occurrence matrices without accessing semantic foundations that have traditionally distinguished human cognitive processing (Dupre, 2021).

### *Training data complications and systemic distortions*

The foundational datasets underlying machine learning architectures introduce systematic distortions that compromise the authenticity of the output by failing to capture global diversity (Hagos *et al.*, 2024). Given that computational models extract correlational structures from input corpora, any demographic imbalances, gaps, or selective sampling within these datasets become embedded within the resulting systems. Consider scenarios where language architectures receive training primarily from materials authored by narrow population segments; such models demonstrate reduced proficiency when processing linguistic varieties from underrepresented communities.

These limitations manifest as reinforcement of social prejudices, systematic exclusion of minority viewpoints, or inadequate recognition of geographical and cultural linguistic diversity. Additionally, the scale requirements of contemporary machine learning necessitate largely unmoderated data collection, inevitably incorporating misinformation and factual inaccuracies that propagate through system outputs (Zyda, 2024).

When learning algorithms process corrupted information, they subsequently reproduce and magnify these distortions. For instance, architectures exposed to conspiracy-laden or prejudiced publications may subsequently generate content promoting such problematic narratives. Moreover, given language’s inherent context-dependency, identical expressions carry vastly different meanings across speakers and situations. Computational systems exhibit fundamental limitations in contextual interpretation, frequently leading to informational misanalysis (Dupre, 2021). Consequently, despite careful development practices, inaccurate content infiltrates system outputs, creating systematic distortions.

## Conclusion

This investigation explored fundamental constraints affecting artificial intelligence systems when processing and resolving ambiguous linguistic phenomena, particularly emphasizing difficulties encountered with Turkish, a language characterized by rich morphological complexity and syntactic flexibility.

This study directly addresses a critical gap in LLM research by shifting its focus to morphologically complex languages, such as Turkish, which are often underrepresented in training data. While previous work has generally noted AI's struggle with low-resource languages, the findings provide a diagnostic challenge by demonstrating that current models consistently fail to resolve ambiguities that rely on deep morphological or scope-based structural understanding. The errors generated by the AI are not just translation mistakes; they reveal a fundamental weakness in integrating surface-level statistical patterns with the grammatical competence necessary for accurate disambiguation in Turkish, suggesting that AI's reliance on large-scale performance data is insufficient when parsing high-entropy languages.

Although AI technologies, primarily through transformer-based language models, have transformed numerous linguistic applications, their dependence on statistical regularities extracted from observable language use creates inherent barriers to accessing more profound grammatical knowledge. This limitation proves especially pronounced in Turkish, where morphological fusion, inflectional marking, and variable constituent ordering generate interpretive challenges that current computational approaches cannot resolve without human-level contextual reasoning.

Dupre (2021) conceptual framework identifies a fundamental divide between surface-level pattern detection (characteristic of artificial systems) and underlying grammatical competence (unique to human cognition). Computational programs may achieve increasingly sophisticated language simulation capabilities without developing the abstract cognitive structures necessary for genuine linguistic comprehension. McShane & Nirenburg (2021) similarly examine multiple dimensions of linguistic uncertainty and the constrained analytical capacity of current AI technologies. The Turkish case studies presented illustrate how AI's propensity for generating structurally correct yet semantically problematic or erroneous outputs, commonly called "synthetic fabrications", derives from this surface-oriented methodology.

Contemporary AI represents what scholars classify as "narrow" intelligence, designed to replicate human mental processes rather than "general" intelligence that would create authentic cognition capable of genuine understanding and mental state representation (Flowers, 2019; Hockly, 2023). This reality emphasizes the critical need for human supervision and analytical oversight. Nevertheless, while current AI faces significant constraints, the eventual achievement of general artificial intelligence could, as Dupre (2021) acknowledges, enable AI systems to perform linguistic analysis comparable to that of humans, allowing researchers to examine their processing mechanisms and potentially revolutionizing cognitive science research methodologies. Sindhu *et al.* (2024) identify emerging AI developments focused on extended contextual processing, bias reduction, and computational optimization through distributed learning approaches. Therefore, numerous unforeseen AI capabilities remain on the horizon.

Looking ahead, enabling AI to effectively handle linguistically ambiguous languages like Turkish and genuinely close the divide between artificial and human language processing requires a fundamental theoretical restructuring of AI architectures to incorporate comprehensive linguistic knowledge. This transformation demands progression beyond simple correlation matching toward authentic language comprehension, necessitating mastery of grammatical principles and semantic nuances that transcend statistical associations.

Developing AI expertise and promoting collaborative efforts between linguistic researchers and AI engineers will prove crucial for creating robust systems capable of managing complex languages, such as Turkish, while advancing linguistic research. By maintaining a balance between technological advancements and rigorous evaluation, we can empower AI to contribute to linguistic knowledge and enhance language education globally.

## Suggestions for further studies

The exploratory nature of this investigation, which focused on a limited corpus of ten diagnostic Turkish sentences across five distinct AI models, lays a necessary groundwork but inherently points toward several avenues for comprehensive future research.

First, the primary limitation of this study was its scope and lack of a human-AI comparison. Future research can conduct a rigorous human baseline experiment where a group of native Turkish speakers disambiguate the exact same ten sentences, both with and without context. This would allow for a direct, statistically evaluated comparison between established human parsing preferences and the AI's disambiguation choices.

Second, while this study provided qualitative insights into the five selected models, a proper statistical evaluation of their performance is required. Future work can quantify which AI models or model architectures (e.g., encoder-decoder, decoder-only) are significantly more accurate, faster, or more consistent in handling specific types of ambiguity (e.g., Lexical, Morphological, Structural) in Turkish.

Third, the testing corpus could be expanded to increase generalizability. Using a larger dataset of hundreds of sentences per ambiguity type would move the findings beyond this diagnostic set, enabling researchers to determine if the observed AI behaviors are a general feature of the models or merely an artifact of the small sample size.

Finally, given that the most challenging ambiguities are often Structural (like Quantifier Scope and Relative Clause Attachment, as seen in the selected sentences), future studies might focus on testing the AI's ability to utilize linguistic context and semantic plausibility to resolve these specific structural conflicts, potentially through iterative prompting techniques that mirror the dynamic, constraint-based nature of the human parsing process.

**Declaration of generative AI and AI-assisted technologies in the writing process:** During the preparation of this work, the author used Grammarly, Gemini, and Claude.ai to improve language and readability of the text. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## Funding statement

This research received no external funding.

## Conflict of interest

The author has no relevant financial or nonfinancial interests to disclose.

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Author contribution

The author confirms being the sole contributor of this work and has approved it for publication. This article constitutes the full-text version of a paper presented at the 22nd International Conference on Turkish Linguistics (ICTL 2025), held on September 4–6, 2025.

## References


1. Alaqlobi, O., Alduais, A., Qasem, F., & Alasmari, M. (2024). Artificial intelligence in applied (linguistics): a content analysis and future prospects. *Cogent Arts & Humanities*, 11(1). https://doi.org/10.1080/23311983.2024.2382422.
2. Backus, A., Cohen, M., Cohn, N., Faber, M., Krahmer, E., Laparle, S., Maier, E., Van Miltenburg, E., Roeflofsen, F., Sciubba, E., Scholman, M., Shterionov, D., Sie, M., Tomas, F., Vanmassenhove, E., Venhuizen, N., & de Vos, C. (2023). Big questions for linguistics in the age of AI. *Linguistics in the Netherlands*, 40, 301–308. https://doi.org/10.1075/avt.00094.bac.
3. Corchado, J. M., López F, S., Núñez V, J. M., Garcia S, R., & Chamoso, P. (https://www.google.com/search?q$=$2023). Generative artificial intelligence: Fundamentals. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 12(1). DOI: https://doi.org/10.14201/adcaij.31704.
4. Curry, N., Baker, P., & Brooks, G. (2024). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*, 4, 1–9. https://doi.org/10.1016/j.acorp.2023.100082.
5. Dupre, G. (2021). (What) can deep learning contribute to theoretical linguistics? *Minds and Machines*, 31, 617–635. https://doi.org/10.1007/s11023-021-09571-w.
6. Flowers, J. C. (2019). Strong and Weak AI: Deweyan Considerations. *AAAI Spring Symposium: Towards Conscious AI Systems* . https://ceur-ws.org/Vol-2287/paper34.pdf.
7. Hadi, M. U., Al-Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Al-Garadi, M. A., Wu, J., & Mirjalili, S. (2023). Large language models: A comprehensive survey of applications, challenges, limitations, and future prospects. https://doi.org/10.36227/techrxiv.23589741.v1.
8. Hagos, D. H., Battle, R., & Rawat, D. B. (2024). Recent advances in generative AI and large language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*, 5(12), 5873–5893. https://doi.org/10.1109/TAI.2024.3444742.
9. Hockly, N. (https://www.google.com/search?q$=$2023). Artificial intelligence in English language teaching: The good, the bad and the ugly. *RELC Journal*, 54(2), 445–451. https://doi.org/10.1177/00336882231168504.
10. McShane, M., & Nirenburg, S. (2021). *Linguistics for the Age of AI*. MIT Press.
11. Prodan, A., Occhipinti, J.-A., Ahlip, R. A., Ujdur, G., Eyre, H. A., Goosen, K., Penza, L., & Heffernan, M. (2024). Cutting through the confusion and hype: Understanding the true potential of generative AI. *SSRN*. https://dx.doi.org/10.2139/ssrn.4995089.
12. Sindhu, B., Prathamesh, R. P., Sameera, M. B., & KumaraSwamy, S. (2024). The evolution of a large language model: Models, applications and challenges. *International Conference on Current Trends in Advanced Computing (ICCTAC)*, 1–8. https://doi.org/10.1109/ICCTAC61556.2024.10581180.
13. Zyda, M. (2024, March). Large language models and generative AI, Oh My! Computer, 127–132. https://://doi.org/10.1109/MC.2024.3350290.