



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

DISTINCTIVE FEATURES OF MULTIMODAL AND LINGUISTIC PROMPTS IN ARTIFICIAL INTELLIGENCE DISCOURSE

Tuopova Marhabo Sadullayevna

English Language Teacher at the Specialized School of Oltinsoy District

E-mail: marhabosadullayevna5@gmail.com,

ORCID ID 0009-0008-3591-7317

Abstract

This article examines the structural, semantic, pragmatic, and discourse features of linguistic and multimodal prompts in artificial intelligence discourse. It identifies the roles of verbal precision, syntactic organization, communicative intention, visual grounding, and cross-modal coherence. The study conceptualizes multimodal prompts as complex discourse units integrating textual and visual information within a shared communicative task, thereby expanding approaches to human–AI communication and contemporary digital discourse research.

Keywords: multimodal prompts, linguistic prompts, artificial intelligence, MLLMs, visual grounding, discourse, communication.

Introduction

The development of artificial intelligence technologies is bringing human communication with digital systems to a new stage. While large language models were initially focused primarily on processing textual information, multimodal large language models have expanded the possibilities for processing text, images, audio, and other forms of information within a unified communicative environment. In such systems, users can control the model not only through



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

written instructions but also through images, graphics, diagrams, or combinations of text and images. As a result, the concept of the prompt has expanded in meaning, becoming a complex communicative unit encompassing both linguistic and multimodal forms. Contemporary research has noted that MLLMs integrate textual and visual modalities and provide dialogue and instruction-following capabilities.

A linguistic prompt expresses a communicative purpose through natural language, whereas a multimodal prompt formulates this purpose through text together with visual or other modal information. For example, in a task such as “Analyze this image,” the linguistic instruction is dominant, while the image itself also becomes a semantic component of the task. Consequently, in multimodal prompts, meaning is generated not only by verbal units but also through the interaction between modalities. The study of this phenomenon therefore requires the syntactic, semantic, pragmatic, and discursive characteristics of prompts to be examined in conjunction with the multimodal context.

Literature Review

The initial directions of prompt research focused primarily on the role of natural-language instructions in controlling the behavior of large language models. Reynolds and McDonell consider prompt programming as a mechanism for directing model behavior through natural language. Liu and co-authors systematize prompting methods and describe the functional capabilities of prompts across various NLP tasks. These studies provide a theoretical basis for explaining the communicative and technological nature of linguistic prompts.

Attention to the direct linguistic characteristics of prompts is further developed in Mohammed's research. Based on an analysis of 250 prompts, the author examines their syntactic, lexical, pragmatic, and discursive characteristics and



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

demonstrates that explicit imperative and interrogative constructions, lexical specificity, and contextual relevance are associated with the quality of AI-generated responses.

Research on multimodal prompts extends the concept of prompting through the incorporation of visual modality. Wang and co-authors systematize visual prompting methods and demonstrate the role of image-based instructions in visual grounding, object identification, and compositional reasoning. Gu and co-authors comparatively examine natural-language and visual forms of prompt engineering in vision-language foundation models. Caffagni and co-authors, in turn, provide a synthesis of the capabilities of multimodal large language models to integrate textual and visual information.

At the same time, a considerable portion of existing research focuses on technical performance, while the linguistic and pragmatic characteristics of multimodal prompts have received comparatively less attention. This situation creates a need for a comparative investigation of linguistic and multimodal prompts within a unified framework of human–AI discourse.

Analysis and Results

In artificial intelligence discourse, the prompt is one of the principal units organizing communicative interaction between the user and the model. In a linguistic prompt, the communicative task is primarily encoded through natural-language units. The user specifies the expected outcome through commands, questions, descriptions, explanations, or constraints. In this process, the syntactic form, lexical precision, and pragmatic purpose of the prompt influence its interpretation. Mohammed's empirical study likewise demonstrates that the precise syntactic structure and lexical specificity of prompts are associated with the accuracy and coherence of AI-generated responses.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

An important characteristic of the linguistic prompt is its expression of communicative intention in verbal form. In constructions such as “Analyze the text,” “Explain this concept,” and “Describe the objects in the image,” the task is explicitly specified through the verb. In such units, the imperative or interrogative form directly encodes the user's communicative purpose. The semantic effectiveness of a prompt, however, depends not only on its grammatical form but also on its context, terminological precision, and the degree to which the task is clearly delimited.

In a multimodal prompt, communicative meaning is formed through the interaction of several modalities. When a user provides an image together with a textual instruction, the image is not merely an illustration but becomes a semantic component of the prompt. In a prompt such as “Analyze the information in this diagram,” “analyze” constitutes the linguistic instruction, while the diagram itself provides the semantic material to be analyzed. Thus, the communicative structure of a multimodal prompt consists of the functional unity of verbal and nonverbal components.

Research on visual prompting demonstrates that images can provide additional information to the model for performing a task and may be more effective than text-only prompts for certain tasks. In particular, visual prompts may contribute to improving the visual grounding and compositional reasoning capabilities of MLLMs. This indicates that the distinction between multimodal and linguistic prompts cannot be adequately explained simply by the “addition of an image.” The distribution of information across modalities and their semantic interrelationships are of central importance.

Referential relations occupy a particularly important position in multimodal prompts. In instructions such as “Explain the first object in the image,” “Analyze the table on the left,” and “Explain the section marked in red,” linguistic units



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

refer to visual objects. In this process, the meaning of expressions such as “the first object,” “the table on the left,” and “the section marked in red” cannot be determined solely from the text itself; their interpretation depends on the visual context. Consequently, deixis, reference, and contextual dependence acquire new forms in multimodal prompt discourse.

In a linguistic prompt, context is constructed primarily through verbal means. The user specifies the topic, audience, purpose, style, and response format in written form. In a multimodal prompt, however, part of the context is automatically provided through an image or another modal component. This leads to a distribution of the communicative burden across modalities. For example, instead of describing a graph in detail verbally, a user can provide the graph itself and specify only the task of analyzing it. In such a case, the image carries a substantial portion of the semantic load.

Another characteristic of multimodal prompts is the issue of correspondence between modalities. When text and image complement each other, the model may interpret the task more accurately. Conversely, if a discrepancy arises between the linguistic instruction and the visual material, communicative ambiguity may occur. Therefore, the effectiveness of a multimodal prompt should not be assessed solely according to the individual quality of each component but also on the basis of their semantic and pragmatic coherence.

Because MLLMs are designed to process textual and visual information jointly, multimodal prompts are directly related to their instruction-following and reasoning capabilities. Caffagni and co-authors demonstrate that the integration of textual and visual modalities is important for tasks such as visual grounding, image understanding, image generation, and other functions of multimodal large language models. Zhang and co-authors systematize the capabilities of



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

multimodal LLMs in performing various multimodal tasks and emphasize the importance of multimodal inputs and outputs in the development of such models. The practical effectiveness of multimodal prompting is also demonstrated by the study of Yoon and co-authors. They converted sensor data into visual representations and presented them to an MLLM as visual prompts. The study evaluated nine sensing tasks across four sensing modalities and reported that visual prompting increased average accuracy compared with text-based prompts while also reducing token costs. This finding suggests that multimodal prompting is not merely a means of increasing the amount of linguistic information but can also serve as a way of structurally adapting information into a modality that is more accessible to the model.

The distinction between linguistic and multimodal prompts can be explained through the distribution of communicative load. In a linguistic prompt, the primary semantic load is conveyed through words, grammatical constructions, and discourse relations. In a multimodal prompt, semantic load is distributed across text, images, and the relationships between them. Therefore, multimodal prompts require users to possess not only linguistic competence but also the competence to select visual material appropriate to the task and connect it effectively with the linguistic instruction.

From a pragmatic perspective, both types of prompts communicate the user's communicative purpose to the model. The difference is that, in a linguistic prompt, the purpose may be expressed more fully in verbal form, whereas in a multimodal prompt some information is assigned to the visual component as contextual information. For example, in the prompt “Write a scientific conclusion based on this image,” “write a scientific conclusion” specifies the pragmatic task, while the image provides the factual context necessary for accomplishing that



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

task. Thus, from a pragmatic perspective, the multimodal prompt functions as a multichannel instruction.

From a discourse perspective, multimodal prompts also expand the boundaries of human–AI communication. Whereas in traditional textual discourse the sequence of meaning is constructed primarily through verbal units, in multimodal AI discourse the relationship between text and image also contributes to the formation of discourse coherence. This makes multimodal cohesion, referential relations, deixis, and semantic integration important new objects of linguistic research.

At the same time, the nature of ambiguity changes in multimodal prompts. In a linguistic prompt, ambiguity is often associated with alternative interpretations of lexical or syntactic units. In a multimodal prompt, ambiguity may arise from inconsistency between text and image, incorrect referential identification of an object in the image, or insufficient connection between the user's instruction and the visual material. Therefore, the analysis of multimodal prompts requires a criterion of cross-modal clarity in addition to linguistic precision.

The results of the analysis also demonstrate that there is no strict boundary between the two forms of prompting. In contemporary AI systems, purely textual prompts may refer to visual elements, while multimodal prompts may contain complex linguistic instructions. Therefore, rather than dividing them into rigid categories such as “textual” and “visual,” it is more scientifically appropriate to classify them according to the relative contribution of verbal and nonverbal components to the communicative task.

In general, the principal characteristics of a linguistic prompt include verbal precision, syntactic organization, lexical specificity, pragmatic purpose, and discourse coherence. In a multimodal prompt, these characteristics are complemented by visual grounding, referential relations, cross-modal



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

correspondence, and visual-semantic context. Therefore, a multimodal prompt can be regarded not simply as an extension of a linguistic prompt but as a complex discursive construction that integrates different information modalities around a unified task in human–AI communication.

Conclusion

In artificial intelligence discourse, linguistic and multimodal prompts are interconnected means of organizing communicative interaction between humans and models. While a linguistic prompt expresses purpose, task, and context through natural language, a multimodal prompt constructs this communicative load through text, images, and their semantic relationships. The findings demonstrate that reference, deixis, visual grounding, cross-modal correspondence, and pragmatic precision are particularly important in multimodal prompts.

The development of contemporary MLLMs increasingly necessitates the study of prompts not merely as textual instructions but as discursive units integrating multiple information modalities. From this perspective, the comparative study of multimodal and linguistic prompts represents a promising direction for identifying new linguistic patterns and regularities in human–AI communication.

References:

1. Fayziyeva A.S. O‘zbek tilida sun’iy intellekt terminlarining shakllanish manbalari va usullari // Qo‘qon DPI. Ilmiy xabarlar jurnali. — 2026. — Vol. 8, № 6. — B. 84–90. — DOI: 10.70728/b.series.fil.v08.i06.010.
2. Odilova S.T. O‘zbek tilida sun’iy intellekt yordamida yaratilgan matnlarning lingvopragmatik xususiyatlari // Yangi Renessansda ilm-fan taraqqiyoti. — 2026. — Vol. 2, № 1.



International Conference on Education, Psychology and Humanities

Hosted Online from Moscow, Russia

Date: 28th September, 2026

Website: <https://econferencia.com>

3. Xolmanova Z.T. Kompyuter lingvistikasi: darslik. — Toshkent: Mehr-nuri, 2024. — 232 b. — ISBN 978-9910-8801-8-6.
4. Abdurahmonova N. Kompyuter lingvistikasi: darslik. — Toshkent: Nodirabegim, 2021. — 394 b. — ISBN 978-9943-6941-2-5.
5. Mohammed G.M.S. A linguistic inquiry into prompt engineering in AI-human communication // Discover Education. — 2026. — Vol. 5. — Article 733. — DOI: 10.1007/s44217-026-01575-x.
6. Reynolds L., McDonell K. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm // Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. — New York: ACM, 2021. — P. 1–7. — DOI: 10.1145/3411763.3451760.
7. Liu P., Yuan W., Fu J., Jiang Z., Hayashi H., Neubig G. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing // ACM Computing Surveys. — 2023. — Vol. 55, No. 9. — P. 1–35.