American Journal of Social and Humanitarian Research

Vol. 6 Issue 9 | **pp.** 2270-2274 | **ISSN:** 2690-9626
Available online @ https://globalresearchnetwork.us/index.php/ajshr

Article

Semantic Activation in Uzbek: A Corpus-Based Model of Potentiality, Virtuality and Actuality

Ziyadullayeva Madina[*1]

1. PhD student at Samarkand State Institute of Foreign Languages
* Correspondence: tursunboyevamadina98@gmail.com

Abstract: This study introduces a corpus-based semantic model for analyzing linguistic units in Uzbek, grounded in the triadic framework of potentiality, virtuality, and actuality. Drawing from cognitive linguistics and discourse theory, the model captures how lexical items transition from latent existence to contextual activation in real speech. Through annotated text analysis and semantic energy mapping, the paper identifies linguistic indicators of activation and proposes a methodology for computational annotation. The findings offer a foundation for developing annotated corpora and AI-based language models for Uzbek, a low-resource language.

Keywords: Semantic Activation, Uzbek Corpus, Potentiality, Virtuality, Actuality, Annotation, Discourse Modeling, Cognitive Linguistics, Discourse Analysis

Citation: Madina Z. Semantic Activation in Uzbek: A Corpus-Based Model of Potentiality, Virtuality and Actuality. American Journal of Social and Humanitarian Research 2025, 6(9), 2270-2274.

Received: 30th Jul 2025
Revised: 10th Aug 2025
Accepted: 30th Aug 2025
Published: 23rd Sep 2025

1. Introduction

Semantic activation—the process by which linguistic units acquire contextual meaning—is central to both theoretical linguistics and corpus-based analysis. In Uzbek linguistics, this phenomenon is traditionally described as the actualization of potential lexical features in speech (Ne'matov & Muhamedova)[1]. However, existing approaches often treat activation as a binary event, overlooking intermediate states and contextual triggers.

This paper proposes a triadic model—potentiality, virtuality, and actuality—to describe semantic movement in Uzbek discourse[2]. By integrating cognitive linguistics, discourse analysis, and computational annotation, the study aims to provide a scalable framework for corpus tagging and semantic modeling.

2. Materials and Methods

The methodological framework of this research is designed to capture how semantic activation unfolds in Uzbek across three conceptual dimensions: potentiality, virtuality, and actuality. The approach relies on a corpus-based model, which enables the analysis of naturally occurring linguistic data rather than artificially constructed examples[3]. This ensures that the findings reflect authentic language use and the pragmatic functions of meaning in real communicative contexts.

This study adopts a qualitative-quantitative hybrid design. The qualitative dimension allows for the exploration of subtle semantic shifts, metaphorical uses, and culturally embedded expressions, while the quantitative dimension provides statistical

evidence to demonstrate frequency, distribution, and collocational patterns. By combining these approaches, the study balances interpretative depth with empirical reliability.

For this research, the Uzbek National Corpus serves as the primary data source, supplemented by contemporary Uzbek newspapers, literary works, and digital texts (blogs, social media, and academic publications)[4]. This multi-layered corpus is chosen because it: represents both standard literary Uzbek and colloquial variations, covers different genres (prose, poetry, journalism, and online discourse), provides temporal depth, ranging from classical to modern usage. The dataset comprises approximately 10 million words, ensuring both representativeness and manageability.

The phenomenon of semantic activation has long intrigued linguists across traditions. In Uzbek linguistics, it is typically described as the realization of latent lexical features in speech (Ne'matov; Rasulov)[5]. However, this view can be expanded by engaging with broader theoretical frameworks.

Ferdinand de Saussure's distinction between langue and parole laid the groundwork for understanding how linguistic units transition from abstract potential to concrete realization (Saussure)[6]. In this view, potentiality resides within the system, while actuality emerges in speech. Russian structuralists such as Lev Shcherba and Viktor Vinogradov further developed the concept of "потенциальные слова" (potential words) — lexical items that exist in the system but are not yet actualized in discourse (Vinogradov)[7].

G. I. Solntsev, in his systemic-functional approach, emphasized that the meaning of a linguistic unit is not fixed but emerges through it interaction with communicative context. He argued that the semantic structure of a word includes both actualized and non-actualized components, and that activation depends on discourse relevance and speaker intention (Solntsev)[8]. This aligns with the idea that virtuality is a transitional state — semantically summoned but not yet articulated.

V. V. Ivanov, building on semiotic theory, proposed that linguistic signs carry a layered potential that becomes actual only through interpretive engagement. His work on the semantic field and functional readiness of linguistic units supports the notion that meaning is not merely stored, but dynamically activated (Ivanov)[9].

3. Results and Discussion

The corpus analysis revealed a clear distribution of linguistic markers corresponding to the three semantic dimensions: potentiality, virtuality, and actuality. Out of approximately 10 million words, more than 15,000 instances of modal or aspectual constructions were identified and categorized[10].

- Potentiality: Expressions of possibility and capacity ("mumkin", "ehtimol", "qodir") appeared with high frequency in journalistic texts and academic writing, where hypothetical scenarios are often discussed. These markers represented about 42% of the total dataset.
- Virtuality: Constructions related to imagination and unrealized states ("tasavvur qilmoq", "go'yo", "xayolan") were most common in literary prose and poetry, accounting for 28% of the findings. These were strongly linked to metaphorical and figurative usage, often expressing cultural or emotional nuances.
- Actuality: Realized actions and factual states ("amalga oshmoq", "ro'y bermoq", "sodir bo'lmoq") dominated in news reporting and historical accounts, making up 30% of the dataset[11]. These markers were closely tied to event narration and factual documentation.

The data also demonstrated a dynamic transition between categories. In many cases, potential meanings shifted into actuality through virtual or imaginative stages, especially in narrative texts. This suggests that Uzbek speakers tend to construct meaning through gradual activation rather than abrupt shifts.

In Uzbek linguistics, Hamid Ne'matov provided a foundational framework for understanding the relationship between lexical potential and speech realization[12]. He defined actualization as "the process by which a linguistic unit transitions from systemic existence to communicative function," emphasizing the role of syntactic position, morphological emphasis, and pragmatic context. His work laid the groundwork for analyzing activation indicators in Uzbek discourse.

Cognitive linguistics reframes activation as a gradient process. Langacker introduced the concept of activation zones, where linguistic units become salient through attention and context[13]. Evans elaborated on this by linking semantic readiness to mental representation and discourse relevance.

Corpus linguistics, meanwhile, emphasizes empirical observation demonstrated how frequency, co-occurrence, and register variation influence lexical activation. In low-resource languages like Uzbek, annotated corpora are essential for modeling such patterns. Jurafov highlighted the need for semantic tagging in Uzbek NLP, proposing annotation schemes that align with activation theory.

This study synthesizes these perspectives by proposing a triadic model — potentiality, virtuality, and actuality — that captures the full semantic trajectory of a linguistic unit. Unlike binary models, this framework accommodates transitional states and contextual triggers, making it suitable for both theoretical analysis and computational implementation[14].

The relationship between potentiality, virtuality, and actuality in Uzbek semantics can be further clarified through a three-dimensional energy model. This model conceptualizes semantic states not only as abstract categories but also as graded levels of activation, where potentiality represents the passive lexical base, virtuality embodies predictive readiness, and actuality signals full realization within discourse.

Table 1. Semantic Activation States in Uzbek

State	Semantic Energy	Discourse Presence	Computational Equivalent
Potentiality	0 (passive)	Absent	Lexical database entry
Virtuality	1 (ready)	Predicted	Contextual prediction
Actuality	2 (active)	Present	Token + semantic tag

As shown in Table 1, each state reflects a distinct level of semantic energy. Potentiality exists in the background as a passive lexical entry, virtuality emerges as a predictive stage where meaning is anticipated but not yet realized, and actuality becomes active within discourse as a realized token with full semantic tagging. This tiered framework allows us to analyze how Uzbek discourse moves fluidly between states of possibility, imagination, and factual realization.

Beyond the three-dimensional model of semantic activation, Uzbek also employs a range of linguistic indicators that signal how meaning becomes foregrounded in discourse. These indicators can be observed at different levels of linguistic organization, from syntax to discourse markers.

Table 2. Linguistic Indicators of Semantic Activation in Uzbek

Indicator Type	Example	Function
Syntactic	Bugun men emas, u keldi. (It wasn't me who came today, it was him.)	Focus via inversion
Morphological	Ugina keldi-ku!(He's the one who came!)	Emphatic suffixes
Pragmatic	Shuni aytmoqchimanki...	Speaker intention

Discursive	(What I'm trying to say is...) Shuning uchun u kechikdi. (That's why he was late.)	Causal connector

As shown in Table 2, semantic activation in Uzbek is not limited to modal or lexical choices but extends across multiple linguistic domains. Syntactic structures such as inversion highlight focus, morphological markers strengthen emphasis, pragmatic phrases foreground speaker intention, and discursive connectors establish logical relationships. Together, these indicators illustrate how potential meanings are strategically activated and realized in actual communication.

Example: Bugun men emas, u keldi.

- Bugun → actual (temporal anchor)
- men → virtual (contrastive subject)
- u keldi → actual (predicate)

Example: Shuni aytmoqchimanki, bu ishni faqat siz bajara olasiz.

- Shuni aytmoqchimanki → pragmatic trigger
- bu ishni → virtual object
- siz bajara olasiz → actual predicate

The triadic model reveals that semantic activation in Uzbek is not a binary switch but a gradual transition shaped by context, syntax, and speaker intention. Annotated corpus analysis confirms that activation indicators are consistent across genres and registers. This model can inform NLP applications, especially in low-resource language settings where semantic nuance is critical.

The model supports:

- Corpus annotation: Tagging units by activation level
- AI training: Teaching models to recognize semantic readiness
- Feature extraction: Identifying activation triggers
- Tokenization strategies: Segmenting text based on semantic energy[15]

These applications are vital for developing Uzbek-language NLP tools, including chatbots, translation systems, and semantic search engines.

Table 3. Stages and Tools of Semantic Activation in Uzbek-Language NLP.

Semantic activation	**Transition from lexical existence to contextual meaning**
Potentiality	Lexical presence without discourse use
Virtuality	Contextual readiness without articulation
Actuality	Active use in discourse
Annotation	Semantic tagging of linguistic units
Tokenization	Segmenting text into analyzable units
Semantic activation	Transition from lexical existence to contextual meaning

As illustrated in Table 3, the process of semantic activation is closely connected with the transition from potential lexical existence to actual discourse meaning. The stages of potentiality, virtuality, and actuality trace this continuum, while annotation and tokenization function as the methodological tools that enable such processes to be systematically captured and analyzed.

4. Conclusion

This study presents a corpus-based semantic model for Uzbek, grounded in the triadic framework of potentiality, virtuality, and actuality. By formalizing activation stages

and linking them to linguistic indicators, the model offers a robust tool for corpus annotation and computational linguistics. Future research may explore dialectal variation and expand annotated corpora for Uzbek NLP development. This study demonstrates that semantic activation in Uzbek operates dynamically across potentiality, virtuality, and actuality, reflecting both linguistic structure and cultural cognition. The findings highlight the importance of a corpus-based approach for understanding meaning-making processes in Turkic languages.

REFERENCES

[1] H. Ne'matov, Tilshunoslik nazariyasi. Toshkent: Fan, 2005.

[2] F. de. Saussure, Course in General Linguistics. Duckworth, 1983.

[3] D. Biber, S. Conrad, и D. Reppen, Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, 1998.

[4] R. W. Langacker, Cognitive Grammar: A Basic Introduction. Oxford University Press, 2008.

[5] V. Evans, A Glossary of Cognitive Linguistics. Edinburgh University Press, 2007.

[6] G. I. Solntsev, Osnovy teorii yazyka [Foundations of language theory]. Moscow: Vysshaya Shkola, 1977.

[7] V. V. Vinogradov, O yazyke khudozhestvennoy literatury [On the language of literary texts]. Moscow: Nauka, 1980.

[8] D. Karimova, Evidensiallikning o'zbek va qozoq tillaridagi ifodalanishi. Toshkent: Fan va Texnologiya, 2018.

[9] A. Jurafov, Kompyuter lingvistikasi asoslari. Toshkent: Innovatsiya, 2021.

[10] S. Aliqulov, Modallikning grammatik, semantik va pragmatik tuzilmalari. Buxoro: BuxDU Ilmiy Nashriyoti, 2022.

[11] M. Xudoyberganova, O'zbek tilida zamon, aspekt va modallik tizimi. Toshkent: Ma'naviyat, 2019.

[12] A. Agostini, T. Usmanov, U. Khamdamov, N. Abdurakhmonova, и M. Mamasaidov, UzWordNet: A Lexical-Semantic Database for the Uzbek Language. Tashkent: Innovatsiya Nashriyoti, 2021.

[13] L. V. Shcherba, Yazykovaya sistema i rechevaya deyatelnost' [Language system and speech activity]. Leningrad: Nauka, 1974.

[14] V. V. Ivanov, Znak, sistema, kommunikatsiya [Sign, system, communication]. Moscow: Nauka, 1981.

[15] I. Raxmonov, Frazeologik birliklarning semantik xususiyatlari: Ingliz va o'zbek tillari taqqosida. Samarqand: SamDU nashriyoti, 2020.