THE USE OF DISCOURSE MARKERS IN INFORMAL SPOKEN ENGLISH: A YOUTUBE-BASED CORPUS STUDY AMONG UZBEK ENGLISH SPEAKERS

Andijan region

Pakhtaobod district

English Language Teacher of School 20

Komilova Mukhlisakhon

muxlisakomilova2003@gmail.com

+998945672080

Abstract: This study examines the role and frequency of discourse markers in informal spoken English produced by Uzbek English speakers. The research is based on a corpus of more than 50 YouTube videos, including foreign language learning channels, personal lifestyle vlogs, conversational interviews, and spontaneous speaking content. The total dataset consists of manually transcribed speech, which was carefully analyzed to identify recurring discourse markers and their pragmatic functions. The findings indicate that because, and, like, just, but, and so are the most frequently used discourse markers in informal spoken English. These markers are not only grammatical connectors but also pragmatic tools that regulate discourse flow, manage hesitation, express stance, and enhance conversational naturalness. The study highlights that YouTube-based spoken data provides a rich source for analyzing authentic language use in digital environments.

Keywords: discourse markers, informal speech, corpus linguistics, YouTube data, pragmatics, stylistics

INTRODUCTION: In modern linguistics, spoken discourse has gained increasing attention due to the expansion of digital communication platforms. Among these platforms, YouTube has become one of the most influential sources of authentic spoken English.

Unlike traditional written corpora, YouTube videos contain spontaneous speech, natural interaction, and real-time communication strategies used by speakers in informal contexts. Informal spoken English is characterized by spontaneity, lack of strict grammatical planning, and high dependence on pragmatic elements such as discourse markers[1]. These markers serve as essential tools that help speakers organize their thoughts, maintain fluency, and interact effectively with listeners. In many cases, they replace complex syntactic structures and allow speakers to communicate more naturally[2]. The present study focuses on Uzbek English speakers who produce content in English on YouTube, including educational channels, personal blogs, and conversational videos. The main objective is to analyze how discourse markers function in their informal speech and how these markers contribute to stylistic and pragmatic meaning construction.

THEORETICAL FRAMEWORK: Discourse markers have been widely studied within discourse analysis, pragmatics, and conversation analysis. Schiffrin defines discourse markers as sequentially dependent elements that bracket units of talk and signal relationships between discourse segments[3]. They do not contribute directly to propositional meaning but play a crucial role in coherence and interactional organization. Fraser expands this definition by categorizing discourse markers into functional groups, including:

Additive markers (and, also)

Contrastive markers (but, however)

Causal markers (because, so)

Elaborative markers (like, I mean)[4]

[1] McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

[2] James Paul Gee J. P. An Introduction to Discourse Analysis: Theory and Method. — London: Routledge, 2014.

[3] Schiffrin D. Discourse Markers. — Cambridge: Cambridge University Press, 1987.

[4] Fraser B. What are discourse markers? // Journal of Pragmatics. — 1999. — Vol. 31. — P. 931–952.

From a stylistic perspective, discourse markers are especially important in spoken language because they reflect speaker attitude, hesitation, emphasis, and interpersonal engagement. In informal speech, they often lose strict grammatical boundaries and become multifunctional pragmatic devices. Recent corpus-based studies show that digital spoken platforms such as YouTube, TikTok, and podcasts provide authentic linguistic data that reflect real-world usage rather than prescriptive grammar rules[5]. This makes them highly valuable for studying informal discourse structures.

METHODOLOGY: This research employs a mixed-method corpus-based approach combining qualitative and quantitative analysis. The dataset consists of 50+ YouTube videos selected based on their relevance to informal spoken English. The videos were categorized into three main groups:

1. Foreign language learning channels – including English teaching videos, speaking practice sessions, and pronunciation guides

2. Personal vlog content – daily life recordings, travel experiences, storytelling, and opinion-based videos

3. Conversational and interview-style videos – informal dialogues, Q&A sessions, and spontaneous discussions

All selected videos were transcribed manually to ensure accuracy in capturing spoken features such as fillers, pauses, repetitions, and discourse markers. Unlike automated transcription tools, manual transcription allowed for more precise identification of pragmatic elements.

After transcription, discourse markers were extracted and coded based on:

Frequency of occurrence

[5] McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

Functional category (causal, additive, contrastive, etc.)

Position within utterances

Contextual usage in speech flow

The analysis focused on identifying patterns in how Uzbek English speakers use discourse markers in natural spoken interaction.

RESULTS: The analysis revealed that discourse markers are highly frequent and structurally significant in informal YouTube speech. The most dominant markers across the dataset were and, because, but, so, like, and just. The marker ‘and’ appeared as the most frequent discourse connector. It is used not only to join clauses but also to extend speech without strict syntactic planning. In many cases, speakers repeatedly use ‘and’ to maintain fluency and avoid pauses, which is a typical feature of spontaneous speech production[6]. Moreover, observational evidence suggests that the marker “and” may, in some cases, become overgeneralized to the extent that it functions as a habitual or quasi-filler element in the speech of certain Uzbek English speakers. In particular, instances were noted where “and” was used with exceptionally high frequency, at times resembling what is commonly described in discourse studies as a “filler-like” or “parasitic” usage. For example, in one recorded sample of approximately thirty minutes of spontaneous speech, the marker “and” was employed more than thirty times, often serving not to establish clear semantic coordination, but rather to sustain fluency and avoid silence. The marker ‘because’ is widely used to express causal relations. In personal vlog narratives, speakers frequently justify actions, decisions, or emotions using this marker. It often introduces extended explanations, reflecting a tendency toward self-disclosure and personal reasoning[7]. The marker ‘but’ functions as a contrastive device, allowing speakers to shift perspectives, correct themselves, or introduce alternative viewpoints. It plays an important role in maintaining dynamic interaction in speech. Furthermore, it can be observed that

[6] McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

[7] Guy Cook G. Discourse. — Oxford: Oxford University Press, 1989.

Uzbek learners of English tend to rely heavily on the marker “but” as a primary contrastive device, particularly at lower levels of proficiency, where the range of discourse markers is relatively limited. In such cases, “but” often serves as the default option for expressing opposition or correction. However, this preference is not restricted to lower-level learners alone. Even among speakers with advanced (C1) proficiency, “but” remains one of the most frequently used markers in spoken discourse. This widespread use may be explained by its high functional load, cognitive accessibility, and strong entrenchment in both instructional materials and authentic input. Consequently, the persistent dominance of “but” across proficiency levels highlights its central role in structuring contrastive relations, while also suggesting a potential underutilization of more varied or nuanced alternatives such as “however,” “nevertheless,” or “on the other hand[8].” The marker ‘so’ is commonly used as a discourse transition device. It signals conclusions, results, or topic shifts[9]. In many instances, it appears at the beginning of utterances, indicating spoken rather than written structure. The marker ‘like’ is one of the most multifunctional elements. It serves as a filler, approximation marker, example introducer, and hesitation device. In informal speech, it contributes significantly to the naturalness of spoken English and reflects real-time thinking processes[10]. In addition, it is worth noting that the marker “like” appears to be particularly prevalent among undergraduate students majoring in English as a foreign language at institutions such as the Uzbekistan State World Languages University and the Andijan State Institute of Foreign Languages. Based on informal observational data collected within my own peer group, this discourse marker is among the most frequently employed in spontaneous speech. Such a tendency may be attributed to learners’ exposure to authentic English input through media, as well as their attempt to emulate native-like fluency and manage real-time speech production. Therefore, the high frequency of “like” in this context not only reflects its multifunctionality but also highlights its role

[8] Fraser B. What are discourse markers? // Journal of Pragmatics. — 1999. — Vol. 31. — P. 931–952.

[9] James Paul Gee J. P. An Introduction to Discourse Analysis: Theory and Method. — London: Routledge, 2014.

[10] Deborah Tannen D. Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse. — Cambridge: Cambridge University Press, 2007.

as a pragmatic tool in the developing interlanguage of EFL learners. The marker 'just' functions as a softener and intensifier depending on context. It reduces the force of statements, making speech more polite, casual, and less categorical[11].

The stylistic analysis reveals several key patterns:

High frequency repetition: especially with and and like, indicating spontaneous speech planning

Reduced syntactic complexity: discourse markers replace formal grammatical structures

Interactional fluency: markers help maintain conversation flow without pauses

Hedging and softening: frequent use of just and like reduces assertiveness

Informality dominance: speech is highly conversational and listener-oriented

These features collectively demonstrate that discourse markers are essential for maintaining spoken fluency in informal digital communication.

DISCUSSION: The findings suggest that Uzbek English speakers in YouTube-based informal speech largely align with global English-speaking patterns. The frequent use of like and just reflects native-like tendencies observed in spoken English corpora such as the British National Corpus and American conversational datasets[12]. YouTube as a linguistic data source provides unique advantages: it captures natural speech, allows access to diverse speakers, and reflects real-time communication strategies. However, it also shows variability in language accuracy, as informal speech often includes redundancy, repetition, and non-standard structures. Despite this, discourse markers should not be viewed as linguistic errors. Instead, they function as essential pragmatic tools that enhance communication efficiency and social interaction. Additionally, data obtained from a small-

[11] McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

[12] McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

scale survey conducted among undergraduate students majoring in English language studies further supports these findings. The respondents emphasized that discourse markers not only contribute to the naturalness and fluency of speech but also serve an important cognitive function. In particular, they noted that such markers provide speakers with brief planning time, enabling them to organize their thoughts and formulate subsequent utterances more effectively. This perspective aligns with existing research in discourse analysis, which highlights the role of discourse markers as facilitators of real-time language processing[13]. Therefore, beyond their pragmatic and interactional value, discourse markers can also be understood as strategic devices that assist learners in managing the cognitive demands of spontaneous speech production.

CONCLUSION: This study confirms that discourse markers are fundamental components of informal spoken English. Based on the analysis of YouTube video transcripts, the most frequently used markers among Uzbek English speakers are because, and, like, just, but, and so. These markers perform multiple linguistic and pragmatic functions, including:

Structuring discourse logically

Maintaining fluency in spontaneous speech

Expressing speaker attitude and stance

Enhancing conversational naturalness

Overall, discourse markers represent an essential aspect of spoken language competence. Future research may focus on comparing native and non-native speaker usage or analyzing differences between formal and informal digital communication contexts. Additionally, the findings indicate that the distribution and overreliance on certain discourse markers may also reflect the developmental stage of learners' interlanguage system. In particular, the frequent use of a limited set of high-frequency markers such as

[13] Fraser B. What are discourse markers? // Journal of Pragmatics. — 1999. — Vol. 31. — P. 931–952.

and, but, and like suggests that speakers tend to prioritize cognitive ease and communicative efficiency over lexical diversity in real-time production. This phenomenon is closely associated with spoken language processing constraints, where learners rely on familiar pragmatic resources to reduce cognitive load and maintain speech continuity. Therefore, discourse marker usage not only reflects pragmatic competence but also provides insight into underlying cognitive mechanisms involved in second language speech production. Furthermore, the results highlight the pedagogical implications of discourse marker awareness in English language teaching. Since discourse markers play a crucial role in achieving fluency and naturalness, explicit instruction in their varied functions could help learners expand their repertoire beyond a small set of commonly used items. Classroom activities focusing on authentic spoken input, such as YouTube-based materials, role-plays, and interactive discussions, may enhance learners' ability to use a wider range of discourse markers appropriately. This, in turn, would contribute to more sophisticated and native-like spoken performance, particularly in academic and professional communication contexts.

References

Cook G. Discourse. — Oxford: Oxford University Press, 1989.

Fraser B. What are discourse markers? // Journal of Pragmatics. — 1999. — Vol. 31. — P. 931–952.

Gee J. P. An Introduction to Discourse Analysis: Theory and Method. — London: Routledge, 2014.

McCarthy M. Discourse Analysis for Language Teachers. — Cambridge: Cambridge University Press, 1991.

Schiffrin D. Discourse Markers. — Cambridge: Cambridge University Press, 1987.

Tannen D. Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse. — Cambridge: Cambridge University Press, 2007.

van Dijk T. A. Discourse and Context: A Sociocognitive Approach. — Cambridge: Cambridge University Press, 2008.