VOLUME-4, ISSUE-8

CORPUS-BASED CONTRASTIVE ANALYSIS OF PRAGMATIC MEANING IN UZBEK AND ENGLISH ENTREPRENEURIAL DISCOURSE

Shukhratova Mukhlisa
Basic doctoral student (PhD candidate)
Kokand State University, Uzbekistan

Abstract. The paper outlines a methodology for the corpus-based contrastive study of Uzbek and English entrepreneurial discourse and applies it to the analysis of pragmatic meaning. It distinguishes comparable and parallel corpora, formulates criteria for balancing material by genre, channel, communicative purpose, participant roles, period, origin and size, and proposes a multi-layer annotation scheme. Six layers of pragmatic meaning are described: propositional content, presupposition, implicature, illocutionary force, axiological meaning and genre context. The study shows that communicative functions are isomorphic in the two languages, while their grammatical coding is allomorphic, and that automatic search can only pre-select candidate units, leaving pragmatic interpretation to contextual expert verification.

Keywords: *entrepreneurial discourse, comparable corpus, parallel corpus, annotation, functional equivalence, presupposition, implicature, speech act, isomorphism, allomorphism, Uzbek, English.*

Corpus methods have become a standard tool in contrastive linguistics, yet their application to pragmatic phenomena remains problematic. Frequencies of words and grammatical forms can be counted automatically, but the communicative force of an utterance – whether it makes an offer, softens a refusal or presents a decision as legitimate – depends on context. This problem is especially acute when comparing typologically different languages such as Uzbek, an agglutinative language in which grammatical relations are largely expressed by affixes, and English, an analytic language that relies on auxiliary verbs, conjunctions and fixed word order. Entrepreneurial discourse, which is highly goal-oriented and rich in indirect speech acts, is a convenient field for testing a methodology that combines corpus evidence with pragmatic interpretation.

The theoretical background of the study comprises corpus design principles formulated by D. Biber, S. Conrad and R. Reppen and by T. McEnery and A. Hardie [2; 8], the contrastive analysis framework of C. James [6], speech act theory [1] and H.P. Grice's theory of implicature [5]. Research on Uzbek corpora by M. Sharipov and colleagues, by B. Elov, M. Primova and M. Amirkulov, and by M. Samatboyeva highlights the role of morphological tagging, script normalisation and manual verification [11; 4; 10]. The aim of this paper is to propose and illustrate a corpus-based procedure for the contrastive description of pragmatic meaning in Uzbek and English entrepreneurial discourse.

Comparable and parallel corpora

Two types of empirical sources must be distinguished. A **comparable corpus** consists of texts produced independently in Uzbek and English but matched by genre, communicative purpose, channel and participant roles: an Uzbek *tijorat taklifi* is compared with an English *commercial offer*, an Uzbek startup presentation with an English *startup pitch*. A **parallel corpus** pairs source texts with their translations and serves to test whether meaning, pragmatic force and structure are preserved or shifted

in translation. Since translated texts may carry interference from the source language, mixing them with original texts distorts the comparison. Original texts therefore form the main layer of analysis, while translations constitute a separate control layer.

Balancing criteria and corpus metadata

Reliable comparison requires the material to be balanced according to several criteria. First, **genre**: if one subcorpus consisted mainly of advertisements and the other of contracts, the differences found would reflect genre rather than language. Second, **channel**: written e-mails cannot be compared directly with live pitches, as ellipsis and dialogicity vary across channels. Third, **communicative purpose** – persuading, negotiating, directing or informing. Fourth, **participant roles**: an entrepreneur addressing an investor, a company addressing a customer and a manager instructing an employee occupy different positions, which affects directness, modality and pronoun use. Fifth, **period and size**: texts should come from a comparable time span, and when subcorpora differ in size, frequencies must be normalised per thousand words or per hundred sentences. Every text is stored with metadata on source, date, genre, channel, author or organisation, sector, addresser and addressee roles, language, script (Latin or Cyrillic for Uzbek), origin (original or translation) and word count. Analytical sentences constructed by the researcher to illustrate a grammatical contrast are kept separate from authentic data and are not presented as corpus evidence.

Functional equivalence and multi-layer annotation

The unit of comparison is not the form but the function. English *must* is a modal auxiliary, while Uzbek *kerak* is a modal predicative combined with a verbal noun and a possessive marker; nevertheless, *Buyurtmani bugun tasdiqlashingiz kerak* and *You must confirm the order today* are comparable as expressions of obligation. Likewise, the Uzbek conditional suffix *-sa* and the English conjunction *if* differ in form but share a function: *Oldindan to‘lasangiz, 10 foiz chegirma beramiz – If you pay in advance, we will offer a 10% discount*. Indirect requests such as *Yetkazib berish muddatini aniqlashtirib bera olasizmi?* and *Could you clarify the delivery deadline?* receive the same pragmatic tag, while their linguistic realisation is described separately.

Accordingly, annotation proceeds in three layers. The **morphological-lexical layer** records, for Uzbek, conditional, voice, person-number and verbal noun markers within the verb, and for English, modal and auxiliary verbs. The **syntactic layer** records clause relations, voice, conjunctions and word order. The **semantic-pragmatic layer** marks categories such as directive, offer, mitigation, cooperation, persuasion, refusal, commitment and evaluation. The two languages are not forced into one template: shared functional categories are annotated in parallel with language-specific means of expression.

Layers of pragmatic meaning

The analysis identified six interacting layers of pragmatic meaning. (1) **Propositional content**: in *Narxlar keyingi oyda o‘zgarishi mumkin – Prices may change next month*, the core meaning is a probable future change, expressed by Uzbek *-ishi mumkin* and English *may*. (2) **Presupposition**: in *Biz yana hamkorlikni boshladik – We started working together again*, *yana/again* presuppose earlier cooperation; *to‘xtatmoq/stop* and *afsuslanmoq/regret* act as similar triggers. (3) **Implicature**: *Hozircha ushbu shartlarga rozilik bera olmaymiz – We are not able to accept these terms at present* do not explicitly

exclude future agreement, yet in a real negotiation they may function as a softened refusal; the temporal limit *hozircha/at present* leaves the door open. (4) **Illocutionary force**: a question such as *Muddatni aniqlashtira olasizmi?* or *Could you clarify the deadline?* performs a polite request rather than a request for information; commissive force is expressed by *kafolatlaymiz – we guarantee*, where Uzbek encodes the collective subject in the person suffix *-miz* and English uses an overt subject. (5) **Axiological meaning**: units such as *halol, barqaror, samarali* and *fair, sustainable, efficient* evaluate the referent positively, and their value depends on collocation and the decision they legitimate. (6) **Genre and context**: *Hoziroq buyurtma bering – Order now* is a call to action in advertising but may carry different institutional force in an internal instruction.

Pragmatic layers frequently co-occur within a single fragment. In an Uzbek educational text on business planning, the statement that a business idea alone is not enough, followed by *reja tuzish kerak bo‘ladi* (“a plan needs to be drawn up”) and the figurative remark that investors do not give money for “empty words”, combines refutation of a naïve view, normative necessity and implicature. A functional English counterpart – *A business idea alone is not enough. You need a plan based on clear calculations, because investors will not fund an unsupported promise* – expresses necessity through *need to* and causality through *because*. The pragmatic result is isomorphic; the means are allomorphic.

Limits of automatic analysis

Formal markers are only candidates for pragmatic interpretation. Uzbek *mumkin, hozircha, -sa* and English *could, at present, if* can express possibility, a cautious offer, an indirect refusal, a negotiated concession or a conditional benefit. For example, *We could revisit the proposal* may signal a return to compromise after a refusal, a polite postponement at the end of a conversation, or a genuine intention when a new date is being set; Uzbek *Taklifni yana ko‘rib chiqishimiz mumkin* shows the same variability. A two-stage procedure is therefore proposed: automatic or semi-automatic pre-selection of candidate units, followed by contextual verification that takes into account the propositional content, the preceding and following utterances, the addresser–addressee relationship and the actual or expected response. Recent studies on LLM-assisted annotation likewise suggest that automated tools can support, but not replace, expert pragmatic judgement [14].

Conclusion

The corpus-based contrastive study of Uzbek and English entrepreneurial discourse requires distinguishing comparable and parallel corpora, balancing material by genre, channel, purpose, participant roles, period, origin and size, and using functional rather than formal equivalence as the basis of comparison. Obligation, possibility, condition, directivity, commissivity, presupposition, implicature and evaluation are shared communicative needs and thus form an isomorphic layer. Their realisation differs: Uzbek concentrates pragmatic meaning in affixal, person-number, voice and modal-predicative forms, whereas English relies on modal auxiliaries, overt subjects, conjunctions and fixed word order – an allomorphic layer. Corpus technologies provide frequency, distributional and contextual evidence, but quantitative results cannot replace linguistic interpretation. The proposed methodology can be applied in natural language processing for Uzbek, in the development of annotated business corpora and in translation practice.

References

1. Austin J.L. How to Do Things with Words. – Oxford: Clarendon Press, 1962. – 174 p.
2. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. – Cambridge: Cambridge University Press, 1998. – 300 p.
3. Dalieva M. Methods, Challenges, and Ethical Considerations in Data Collection of Corpus Compilation // Innovative Technologica: Methodical Research Journal. – 2024. – Vol. 3, № 3.
4. Elov B., Primova M., Amirkulov M. O‘zbek tili korpusi: tarixiy rivojlanishi, maqsad va vazifalari // Computer Linguistics: Problems, Solutions, Prospects. – 2025. – Vol. 1, № 1.
5. Grice H.P. Logic and Conversation // Syntax and Semantics. Vol. 3: Speech Acts. – New York: Academic Press, 1975. – P. 41–58.
6. James C. Contrastive Analysis. – London: Longman, 1980. – 208 p.
7. Levinson S.C. Pragmatics. – Cambridge: Cambridge University Press, 1983. – 420 p.
8. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2011. – 294 p.
9. Rahimov X. Tabiiy tilni qayta ishlash (NLP)da korpus texnologiyasining o‘rni // Innovatsion tadqiqotlar zamonaviy dunyoda: nazariya va amaliyot. – 2025. – Vol. 4, № 2. – P. 16–21.
10. Samatboyeva M. O‘zbek tili korpusini annotatsiyalash va teglash (umumiy qoida va tizimlar) // Computer Linguistics: Problems, Solutions, Prospects. – 2025. – Vol. 1, № 1.
11. Sharipov M., Mattiev J., Sobirov J., Baltayev R. Creating a Morphological and Syntactic Tagged Corpus for the Uzbek Language // arXiv. – 2022. – arXiv:2210.15234.
12. Searle J.R. Speech Acts: An Essay in the Philosophy of Language. – Cambridge: Cambridge University Press, 1969. – 203 p.
13. Akramjonova M. Parallel matnlar korpusini yaratishning amaliy va nazariy ahamiyati // Fan va innovatsiya. – 2024. – Vol. 2, № 34. – P. 34–37.
14. Yu D., Li L., Su H., Fuoli M. Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis // International Journal of Corpus Linguistics. – 2023. – DOI: 10.1075/ijcl.23087.yu.