

## The Usage Of Corpus In Discourse Research

Boygucheva Lola Urinbaevna

Teacher, Tashkent State University of Uzbek

Language and Literature after Alisher Navoi

[lolaboyguchova@gmail.com](mailto:lolaboyguchova@gmail.com)

**Abstract:** *The article discusses the issue of corpus linguistics for studying discursive phenomena. Despite of modern corpora are widely used in lexical and grammatical studies, their applications, orientation to the research of various discourse phenomena seems to be restricted. The article gives a brief overview of studies and research of verbal irony demonstrated how specialized and general language corpora can be applied in discourse analysis.*

**Key words:** *corpus linguistics, discourse analysis, usage, discursive annotation, verbal irony.*

The problem of choosing research material can be considered one of the most important issue for modern linguistics. It would seem that existing at disposal numerous dictionaries, grammar books, text books, the ability to record oral speech or direct contacts with native speakers, as well as their own intuition, that linguists should not have difficulty with choosing a data source for analysis. However, it should be noted that two significant problems can be faced by researchers the first - each of these sources reflects, only certain aspects of the existence of a language. The second, every source can not provide sufficient information for a specific research. In this sense, it is especially difficult for researchers whose interests are not only individual phenomena related to a particular level of the language system, but also the usage of language in communication, i.e. discourse. One of the difficulties lies in the need to limit artificially the range of external factors that could affect potentially



the speech activity of native speakers. How is this restriction implemented in practice?

The text is the only entity that is given to linguists in direct observation in modern discourse analysis, it is very often used as source of illustrative examples to confirm the theory under discussion [5,20]. Also researchers play important roles as a native speaker in similar important intuition. Intuitive knowledge of the language together with professional knowledge of language allows linguists to construct examples, and potential contexts for their usage. Toward the result of such an attitude, the text is a situation of parallel coexistence of many models and theories that describe the discourse fragmentarily, and it may contradict each other in some cases.

Over the past fifty years reduce the level of subjectivity in the selection and analysis of material in “level” linguistics helps circulation to language buildings. An important advantage of the corpus is the ability to control various “variables” of communication: texts are equipped in the corpus with meta-marking - additional information about the authors, time and place of creation, genre affiliation, etc. The user gets the opportunity for working with texts whose characteristics are relevant with ongoing research. It would seem that the corpus analysis of discursive phenomena, however, for those who engaged in discourse analysis, the possibility of accessing, the corpus is associated with a number of problems. Firstly, the quality of the study is more dependent on the size of corpus and time of its creation. Secondly, there is there is a serious obstacle for linguists engaged in communicative: modern corpuses are focused primarily on the analysis of lexical and/or grammatical phenomena, and units of communication that do not have standard ways of expression (for example, speech acts). As a result, the corpus cannot provide the user with the ability to obtain data on complex units for communications. Therefore, researchers based on their own intuition and reading experience, deny the usage of enclosures in



property studying text and discourse [2,472]. As arguments, thoughts of this point of view are expressed about the impossibility of creating such an annotation, which reflect structural and semantic text properties. Indeed, annotation discursive phenomena remains one of the major problems of corpus linguistics. Besides, researchers doubt the possibility of using statistical analysis in relation to structural and semantic properties of texts.

At first glance, it is impossible to combine the ideology of discourse analysis with the techniques and methods of corpus linguistics. However, it's also known what the analysis of discourse and corpus linguistics, there is a common object of study - consumption, ouzos (as opposed to orientation to system language features inherent in "level" linguistics). Another important factor in favor joining forces is what corpus analysis lets see the facts that turn out to be inaccessible with an "intuitive" approach to the selection and analysis of texts.

An example of the successful application of corpus linguistics in discursive research phenomena can be considered the work of E. Semino and M. Short, who devoted to the ways of presenting speech and thoughts in English texts [6,272]. For this project A corpus which consists of 250 thousand word usage was created. Discursive text annotation, attribution of additional structural information and informative properties of the text are produced manually.

It can be pointed to a number of case studies for oral discourse:, for example the project – "Tales of Dreams" by A.A. Kibrik, V.I. Podlesskaya and others [1,736], which the records of oral narratives (About the dreams of children's dreams) were transcribed and labeled further in terms of the theory of rhetorical text structures [4,243]. Also, rhetorical relations are underlined the corpus research of M. Taboada, which is dedicated coherence and cohesion in dialogic communication [3]. The material for the study of rhetorical relations was parallel with Anglo-Spanish corpus of dialogue recordings that the purpose of agreement on the time of the meeting.



Another example of the usage of the corpus in the study of discourse is the study of anaphoric relationships, is presented in the work of R. Garside, S. Flea Gelstone and S. Botley [7,261]. Studying anaphora is an attempt to answer the question of how, with the help of coreferent groups, semantic connectivity is provided in discourse. Above mentioned that, the first the studies are united by the methodological aspiration of linguists for using enclosures to solve analysis problems of discourse. The second, the desire to answer the question how semantic integrity is ensured, i.e. discourse coherence.

Studying coherence mechanisms is a complex research task for both theoretical and applied value. However, it is necessary to remember that in reality not all communication is perceived by users as a semantically and pragmatically coherent discourse. Violations of the semantic structure of discourse - there is no less interesting to study the phenomenon. In particular, the speaker creates intentionally incoherence, i.e. semantic "inconsistency" between elements of utterance or between utterance and the describing situation which underlies verbal irony. The permanent ways for expressions, available cases are less informative: morphological, syntactic and semantic markup is not enough to detect an irony. So, in the Uzbek National Corpus language can be detected only 6 cases of pointing out the speaker's irony with litters (with irony) in the oral subcorpus:

*Bupti/ unga ko'proq pul bering/ uning ozgina puli bor (said with irony).*

There is a question, does the fact of absence discursive annotations in the national language do impossibility of its usage in the research for making irony? The study of met-pragmatic markers of the mode of communication, including irony as one of modes. National Corpus allows user to form concordances of contexts in ironic intention which is explicitly marked.

From a semiotic point of view, explicit markers of the mode of communication are the indices with participants in the discourse - "link" of the



situation and statement. Their appearance in speech is explained by the need to ensure coherence of discourse in those cases where dual interpretation is possible text / statements. To study met-pragmatic activities for native speakers of the Uzbek language, a concordance was formed of 675 contexts in which the verb is ironic or the prepositional group with irony functions as met-linguistic comments. Analysis concordance has shown that the need for removing illocutionary ambiguity can be caused several reasons:

1) the speaker fears an incorrect interpretation of his intentions by the addressee; in order to avoid misunderstanding, he explicitly indicates his communicative intentions.

2) the speaker wants the addressee for understanding correctly how the statement and reality are related: in the case of bona fi de communication, the statement corresponds to reality, in the case of non-bona fi de communication compliance is violated; this is what is being reported markers.

3) the addressee wants to check whether he / she is correct, to understand the intentions of the interlocutor.

4) the addressee refuses to accept the proposing modus - non-bona fi de communication and brings the dialogue in the source modus.

5) observer (narrator) evaluates statement as ironic, the way of interpreting the text by reader.

The study of verbal irony corpus provides an opportunity to find out what statements classified by native Uzbek speakers with using irony.

Discursive markup allows user to select fragments of texts which containing verbal signals of irony and statistical analysis of the various strategies occurrence and tactics for creating irony in discourse. One of the challenges is faced by user of corpus balancing and representativeness of the corpus. Because, initial objective of our study was to study irony in various areas of communication, moreover, corpus which including recordings of oral

speech, fragments computer-mediated communication and written non-fiction texts in the field of academic and political discourse, publications of various media. Thanks to a variety of sources provides corpus representativeness with regard to balance, the genre diversity of the included texts (for example, for computer-mediated communication, such sources were blogs, forums, social networks, twitter, news feeds and other). On the other hand, balancing the volume of textual material. Volume corpus is about 2.5 million word usage. Information on the composition of the body is summarized in the Table 1.

**Table 1**

Structure of Corpus

	Written texts	Oral texts	Computer-indirect communication
Number of texts	500	500	500
Number of word	774 000	882 000	879 000

The ways of creating irony were divided into three strategies: verbal, discursive (rhetorical critical) and cognitive. These strategies correspond to three groups of factors that influence our perception of discourse.

The first group consists of linguistic factors associated with native speakers explicit and implicit knowledge of conventions, ousos and language norms, as well as with the ability to distinguish normative and acceptable, expecting usage of the language from abnormal, unusual, unexpected.

The second group combines cognitive factors that are associated with the communicants' knowledge of the world and ideas about the normal events.

The third group is discursive (rhetorical) factors that are manifested in the ability to establish semantic relationships between individual segments discourse, including in those corpuses when these segments are not integral

components of a single communicative situation rather divided in time and space.

Discursive markup cannot be performed in the automatic mode because we are interested phenomena associated to not the formal as with the content side of communication. A discursive text annotation is the result of an analytic activity based on contextual and pragmatic components. In addition, the process of creating a scheme is also a process of clarifying existing ideas about how to create irony in utterance and text. Frequency data implementation of the selected strategies are given in the Table 2.

**Table 2**

Frequency of implementation of various irony strategies

	<i>Verbal strategy</i>	<i>Ritoric strategy</i>	<i>Cognitive strategy</i>
<i>Number of contexts</i>	1569	738	591

Corpus studies of irony has proven useful in several respects. The first, it were revealed groups of factors that affecting the possibility of an ironic interpretation. Secondly, it was possible to see something in common that all disparate ways of creating irony unites in discourse – these the properties are the intentional violation of the semantic integrity of the discourse.

## REFERENCES

1. Dream Tales: Corpus Study oral discourse / ed. A. A. Kibrika and VI. Podlesskaya. - M.: Languages of Slavic cultures, 2009 .- 736 p.
2. Fludemik M. Towards a Natural Narratology / M. Fludemik. – London ; New York : Routledge, 2002. – 472 p.





3. Garside R. Discourse Annotation : Anaphoric Relations in Corpora / R. Garside, S. Fligelstone & S. Botley // Corpus Annotation / ed. by R. Garside, G. Leech & T. McEnery. – London ; New York : Routledge, 1997
4. Mann W. Rhetorical Structure Theory : Toward a Functional Theory of Text Organization / W. Mann, S. Thompson // Text. – 1988. – № 8. 243 p.
5. Plungyan V. A. The corpus as a tool and as an ideological logic: about some of the lessons of modern corpus. Linguistics / V. A. Plungyan // Russian language in the scientific lighting. - 2008. - No. 2 (16). - 20p.
6. Semino E. Corpus Stylistics : Speech, Writing and Thought Presentation in a Corpus of English Writing / E. Semino, M. Short. – London ; New York : Routledge, 2004. – 272 p.
7. Taboada M. Building Coherence and Cohesion: Task-oriented Dialogue in English and Spanish / M. Taboada. – Amsterdam ; Philadelphia : John Benjamins, 2004. – 261 p