2. Bender E. M., Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. ACL Anthology.
3. Crystal, D. (2022). Language and the Internet. Cambridge University Press.
4. Dediu D., Levinson S. C. (2021). On the antiquity of language: The reinterpretation of Neanderthal linguistic capacities and its consequences. Frontiers in Psychology.
5. Hale M., Reiss C. (2008). The Phonological Enterprise. Oxford University Press.
6. Ivanov P. (2022). Bilingualism and Language Policy in Russia. Journal of Slavic Linguistics.
7. Smith J. (2023). The Evolution of English in the Digital Age. American Linguistic Review.
8. Tomasello M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press.

COMPONENTAL AND STRUCTURAL ANALYSIS OF THE BNC

Shakhnoza PARPIEVA
Uzbekistan state world languages university
Teacher

Abstract. *This article provides view to the national corpora and their common features. The issue of the emergence and creation of the British National corpus is also being considered. Moreover, the written and oral foundations of the British National corpus are analyzed. Information about BNC Baby and BNC Sampler, which are considered as sub-corpora of the British National corpus, are provided and its features are highlighted.*

Key words: *corpora, the British National Corpus, collection, sub-corpora, spoken, written.*

Presently, a great number of representative world languages corpora including national corpora have been created. The National Corpus of the Russian language, the British National Corpus, the American National Corpus, the Mannheim German Corpus, the French Corpus, the Hungarian National Corpus, the Modern Chinese Corpus may be considered as evidence. The national corpus includes no less than 100 million words, which is a pledge of opportunities for large-scale study of multilevel language units. These are collections of spoken and written texts of different genres, styles, regional and

social variants represented in the language. Moreover, the whole array of texts in the body is systematized, which means that the corpus fixes the order of each word in the sentence in relation to other words, also takes into account its frequency in the given corpus.

Created in 1991-1994 by researchers of Oxford University and Lancaster University, the British National Corpus (BNC) is considered as "the first and best-known national corpus"[112]. The corpus is 100 million word collection and is much larger than its predecessors. The corpus covers British English of the end XX century, represented by a wide variety of genres and is conceived as a sample of typical spoken and written British English.

90% of the British National Corpus consists of samples of written language usage. These examples were taken from regional and national newspapers, scientific journals, periodicals of various scientific directions, fiction books moreover, both published and unpublished materials (such as brochures, letters, student essays, scripts, speeches) and many other sources[113] [Aston & Burnard, 1998].

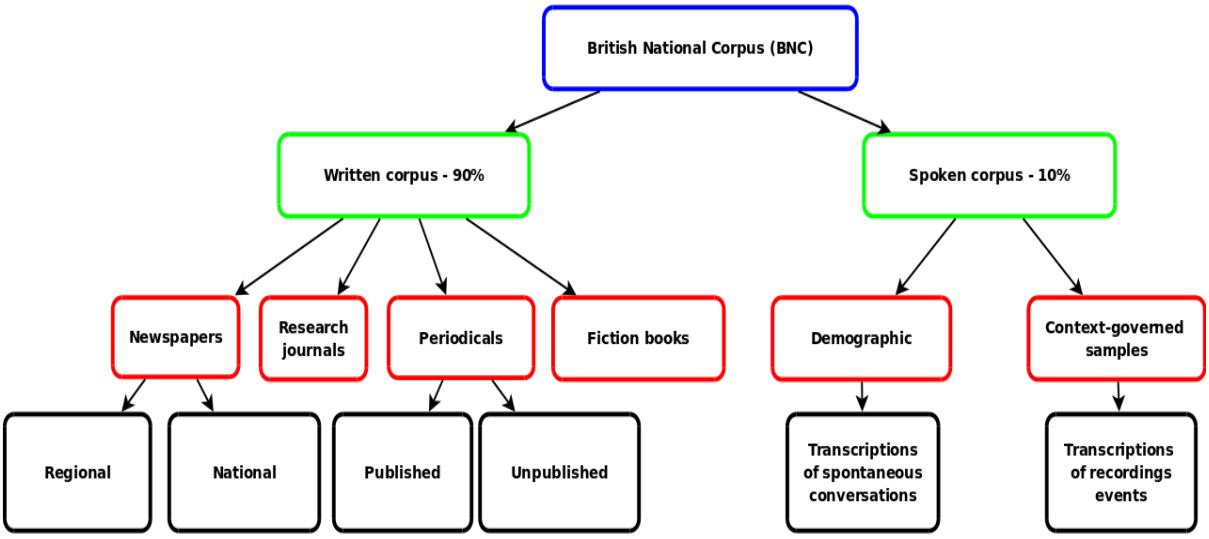
Spoken Corpus holds 10% of the British National Corpus data and examples of conversational language applications, which were presented and recorded using practical transcription. Spoken Corpus includes two large parts: demographic and context-governed. The demographic part contains a transcription of spontaneous conversations that took place in real life with the participation of volunteers from different age groups, regions and social layers. These conversations were produced in a variety of situations, including business or government meetings and discussions in radio broadcasts or by telephone. This was done to take into account both the demographic distribution of the spoken language and the linguistically significant diversity of the language due to the context. The second part of the spoken corpus includes context-governed samples, such as transcriptions of recordings prepared for the entrance of special meetings or events[114].

[112] Xiao, R. Well-known and influential corpora. In: Anke Lüdeling, A. & Kytö, M. (eds.), Corpus Linguistics: An International Handbook. Berlin; New York: Walter de Gruyter GmbH & Co, 2008.
[113] Aston, G. & Burnard, L. The BNC Handbook: Exploring the British National Corpus with SARA. UK: Edinburg University Press, 1998.
[114] Crowdy, S. The BNC spoken corpus. Harlow: Longman, 1995.

Each recording that was included into the spoken corpus is available in the British Library's sound archive and at the Oxford University Phonetics Laboratory website.

Two sub-corpora of the British National Corpus were released under the names BNC Baby and BNC Sampler.

BNC Baby consists of four million word sample and includes four sets of samples. The words in each set correspond to a specific genre category. One set contains transcriptions of conversations, while the other three sets contain samples of written texts from scientific literature, fiction, and newspapers. The subcorpus retains the markup available in the BNC. The newest edition of the BNC was released in XML format.



Two sub-corpora of the British National Corpus were released under the names BNC Baby and BNC Sampler.

BNC Baby consists of four million word sample and includes four sets of samples. The words in each set correspond to a specific genre category. One set contains transcriptions of conversations, while the other three sets contain samples of written texts from scientific literature, fiction, and newspapers. The subcorpus retains the markup available in the BNC. The newest edition of the BNC was released in XML format.

BNC Sampler includes two subcorpora. The first part contains written data, the second part contains spoken language and each part contains one million words. BNC Sampler was originally used to improve the BNC markup process, which eventually led to the publication of BNC World.

References

1. Aston G. & Burnard L. The BNC Handbook: Exploring the British National Corpus with SARA. UK: Edinburg University Press, 1998.
2. Crowdy S. The BNC spoken corpus. Harlow: Longman, 1995.
3. Xiao R. Well-known and influential corpora. In: Anke Lüdeling, A. & Kytö M. (eds.), Corpus Linguistics: An International Handbook. Berlin; New York: Walter de Gruyter GmbH & Co, 2008.
4. Wikipedia, the free encyclopedia. The British National Corpus. // https://en.wikipedia.org/wiki/British_National_Corpus
5. British National Corpus. The British National Corpus User Reference Guide. http://www.natcorp.ox.ac.uk

LINGUISTIC SPECIFICATIONS OF ON-SCREEN SUBTITLES

Oltinoy DAVLATOVA
UZSWLU, ESL teacher

Annotation. *Due to the increased demand for audio visual materials in foreign language as a main source of entertainment, translators had to come up with a more financially stable option to tackle the current issue, which we know as subtitles that prevent them to hire a group of specialists to do the dubbing or voice over. As subtitles have already become quite globally popular, the topic of their being accurate or not also arose. This leads to the linguistic analysis and norms od on-screen subtitles to be set and it will be discussed in this article explicitly.*

Key words: *linguistic, parameter specification, on-screen subtitles, oral speech, information segmentation, audio visual, filtered, source.*

As Davlatova (2024) mentions that with the high demand for the entertainment means that do not require going out, also the growing interest towards foreign movies require an easier, cheaper option content developing and translation. Subtitles could be very handy in tackling this issue since they offer less amount of human involvement compared to dubbing. Diaz Cintas (2012) claims that “As with any other type of translation, subtitles are expected to provide a semantically adequate account of the original dialogue but with the added complication that they must at the same time respect the spatial and temporal specifications…”