



Annotation manual for *ViewBRI* corpus

1. Purpose of the study

Since the introduction of the Belt and Road Initiative (BRI, *yi dai yi lu changyi* 一帶一路倡议) in 2013, it has been promoted as a cooperation platform for countries not only forming the corridor along the historical Silk Road but also in its indirect proximity. As a result, new trade routes leading from and to China have been established to foster the mutual economic growth and the contact between the European countries and China intensified. Considering the fact that Slovakia was one of the first countries which signed the memorandum of understanding in 2015 to express its interest in expanding cooperation with China under the initiative, present study aims at exploring the presentation of the BRI in three Slovak periodicals: *Pravda*, *SME*, and *Denník N*. Due to China's increasing influence in Europe, initially welcomed activities related to the BRI have resulted in a more cautious attitude of some European countries, therefore the *ViewBRI* project investigates the image of the BRI presented in the articles published in the chosen periodicals between 2013 and 2023 to provide qualitative and quantitative evidence on the presentation of the BRI in the Slovak media.

Aims and specific objectives of the *ViewBRI* project

- To analyze the image of the BRI in the Slovak periodicals by making use of discourse analysis and corpus linguistics
- To explore the presentation of the BRI in three Slovak periodicals: *Pravda*, *SME*, and *Denník N*
- To examine the relation between the attitude of the Slovak media towards the BRI and the major topic it relates to
- To investigate how has the media discourse transformed from 2013 to 2023

2. Corpus and data selection

Textual corpora represent useful tools suitable for the language exploration and analysis of different kinds. The textual data comprised in the corpora can be perceived as a valuable source of information which enables the verification of the theoretical basis established for the chosen research questions or to test the falsifiability, completeness, strength, and objectivity of specific hypotheses (cf. Leech 1992; Lemnitzer and Zinsmeister 2015; McEnery and Hardie 2012). Since media discourse is aimed at informing, persuading, and influencing public opinion, the selection of linguistic elements plays an important role in achieving the desired communicative purpose. Taking the production of texts in the form of articles in the periodicals as an example, the purpose is to convey a message, convince, and express ideas or beliefs (cf. Brown and Yule 1984; Štulajterová 2015; Widdowson 2014). Discourse analysis allows for a deeper comprehension of the BRI's image depicted in the Slovak periodical thanks to the identification of the linguistic elements used for its presentation whereas corpus linguistics as a discipline which brings new insights into the study of traditional linguistic research facilitates the combination of quantitative and qualitative methods (cf. Hunston 2002, McEnery and Hardie 2012) for the data analysis and evaluation.

Specialized small-scale Corpus *ViewBRI* comprises articles discussing the BRI which were published in one of the studied Slovak periodicals: *Pravda*, *SME*, and *Denník N*. In order to map the transformation of the BRI's image from a long-term perspective, the articles are collected based on their publication year (2013-2023). Apart from the year of publication, the articles have to include one of the following key words: *Hodvábna cesta* (Silk Road), *projekt novej hodvábnej cesty* (project of New Silk Road), *jeden pás, jedna cesta* (One Belt, One Road). The gathered textual material is examined following the approaches of the discourse analysis so that the linguistic elements depicting the attitude towards the BRI can be identified, categorized, and indexed to facilitate the retrieval and interpretation of the data. Subsequently, the relation between the attitude of the periodicals towards the BRI and the main topic the BRI relates to is determined to scrutinize the image of the BRI from a broader perspective. Lastly, the transformation of the BRI's image in the Slovak media is investigated across the given time span marked by the initial expectations after its launch (2013-2015) gradually changing to a more

cautious attitude towards the BRI (2016-2019) followed by the outbreak of the Covid-19 pandemic in 2020 and the decreased number of promotion activities up to 2023.

3. Annotation categories and tags

Present annotation manual was developed based on the work with the preliminary data consisting of 23 articles published from 2013 to 2015, 1 article, 4 articles, and 18 articles respectively for each year. The articles are selected based on three criteria: discussing the BRI, year of publication (2013–2023), and language mutation (only articles written in Slovak are considered). Each article (text file) is encoded with a set of metadata (see Table 1) to enable a wide variability of filtering options and facilitate its interdisciplinary use in future research.

<doc.id>	Unique id assigned to each article
<doc.year>	Year of publication
<doc.source>	Newspaper name
<doc.title>	Article title
<doc.author>	Name/gender of the author

Table 1. Metadata

Based on the desk research, the preliminary dataset comprising a testing set of articles is identified and processed. A close reading of the textual material is carried out to verify the relevance of the articles taking into consideration the research objectives and to design the annotation tags for the analysis. To provide a comprehensive characterization of the BRI's image in the Slovak periodicals, the variable *year* facilitates the study of the newspaper's attitude towards the BRI across the years whereas the variable *source* is used for addressing the individual differences among the periodicals. Lastly, the variable *title* is used for the investigation of the newspaper's attitude from a macro-perspective by investigating the article's headlines. Remaining variable *author* is not primarily used for the purposes of present study, however, it might be relevant to other research objectives.

Preliminary annotation categories are displayed below:

Attitude	Positive, neutral, negative
Main area	Economics, international cooperation, business, logistics, geopolitics
BRI as core topic	Yes, no

Table 2. Annotation categories

References

- Brown, G. & Yule, G (1984). *Discourse analysis*. Cambridge, Cambridge University Press.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Leech, G. (1992). Corpora and Theories of Linguistic Performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82* (pp. 105-122). Berlin & New York: Mouton de Gruyter.
- Lemnitzer, L. & Zinsmeister, H. (2015). *Korpuslinguistik: Eine Einführung* [Corpus Linguistics: Introduction]. Tübingen: Narr Francke Attempto Verlag.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, Theory and Practice*. New York: Cambridge University Press.
- Štulajterová, A. (2015). "Plays on words in Media Discourse." In *Selected Chapters from Discourse Analysis*. Banská Bystrica: Belianum.
- Widdowson, Henry J (2007). *Discourse Analysis*. Oxford: Oxford University Press.