The Social Sciences 11 (15): 3770-3773, 2016
ISSN: 1818-5800

Corpus Linguistics in Proverbs and Sayings Study: Evidence from Different Languages

Diana Faridovna Khakimzyanova and Enzhe Kharisovna Shamsutdinova
Institute of International Relations, History and Oriental Studies,
Kazan Federal University, Kazan, Russia

Abstract: The study overviews state of the art corpus based approaches to investigation of proverbs and sayings carried out by scholars all over the world. Various corpora provide a broad empirical basis for studying the usage of proverbs and for evaluation of constraints and preferences associated with specific vocabulary. Corpus-based approach provides an opportunity to gain valuable insight about the distribution of proverbs and their role in a language community. The most essential advantage of applying corpora in linguistic researches is its reliable empirical basis, constant updating and availability of proof of using a lexical unit in different meanings from a multiple number of sources of different genres.

Key words: Proverbs, sayings, corpus linguistics, corpus approach to proverbs, Russia

INTRODUCTION

Corpus linguistics has become a popular scientific investigation tool nowadays. The term is now seen as the study of linguistic phenomena through large collections of machine-readable texts: corpora.

Although, the term corpus linguistics first appeared only in the early 1980's, corpus-based language study has a substantial history. The corpus methodology dates back to the pre-Chomskyperiod when it was used by field linguists such as Boas and linguists of the structuralist tradition including Sapir, Newman, Bloomfield and Pike. Although, linguists at that time would have used shoeboxes filled with paper slips rather than computers as a means of data storage and the 'corpora' they used might have been simple collections of written or transcribed texts and thus not representative, their methodology was essentially 'corpus-based' in the sense that it was empirical and based on observed data (McEnery *et al*., 2006).

Thousands of corpora of different languages have been developed for various purposes within project implementations since the first Brown corpus was designed. Among them are the spoken corpus of the Survey of English Dialects, Helsinki Corpus, the International Corpus of English, the corpus of contemporary American English, the Cambridge Learner Corpus, Czech National Corpus, the Callfriend Egyptian-Arabic, the Michigan Corpus of Academic Spoken English, the International Corpus of Learner English, Russian National Corpus, Classical Arabic Corpus, the International Corpus of Arabic, British National Corpus, National Corpus of Polish, Arabi Corpus, the Emirati Arabic Corpus (under construction). They represent general, specialized, written, spoken, synchronic and diachronic and learner corpora where access is available for free or fee-based.

The corpora can be implemented in a number of linguistic investigations ranged from morphological, syntactical analysis to use of corpora in teaching foreign languages, languages for special purposes, translating and studying of idioms, proverbs and sayings which is the subject of our investigation.

MATERIALS AND METHODS

This study is aimed at analyzing approaches to proverbs and sayings study of different languages in order to structuralize investigation in this field. When conducting the research, the articles of Russian and foreign scholars on the subject were thoroughly studied.

Proverbs and sayings study through corpus linguistics: Now a days many research papers and articles are devoted to using corpus linguistic methodology on different purposes. Carrying out of linguistic analysis on up-to-date texts, large database and possibility to use it for a variety of linguistic researches can be considered as advantages. Furthermore, it provides the researchers with the information about the "real" language. So, there are

Corresponding Author: Diana Faridovna Khakimzyanova, Institute of International Relations, History and Oriental Studies, Kazan Federal University, Kazan, Russia

works on application of corpus linguistics in etymological research, using British National corpus to analyze grammar of the spoken english, corpora use in foreign language teaching practice, the use of corpus linguistics in discourse analysis in researching neologisms, idioms, infrequent and marginal language units, teaching professional terms to EFL students using parallel corpus, parallel corpora as a tool for quantitative studies of language-specific lexicon and in researching jokes.

For example, Russian scholar Zavyalova (2013) analyses idiomatic expressions of Russian, Japanese, Chinese and English phraseology with the help of on-line linguistic corpuses. Another Russian scientist Krotova (2016) conducts researches in Semantics of German idioms to find out new meanings of the idioms as well as frequency of meaning and to consider cases of non-idiomatic or weakly idiomatic usage.

The significant scientific research on proverbs and sayings has been carried out by Durco (2014) who investigates one of the problems of empirical paremiology within paremiological experiments finding out of the paremiological invariant «paremiological lemma» and paremiological variants using methods of corpus linguistics. He proposes the methodology for determining the variability of paremiaes and determining the invariant as a suitable candidate for testing of the paremiological core. It is an essential methodological step in reducing paremiological material within paremiological experiments. (Durco, 2014).

Scientist from Matej (2015) examines Slovene paremiology describing the range of proverb variants, their actualizations, transformations and provides examples of non-prototypical usage of proverbs.

Zirker and Winter-Froemel (2015) Renner investigate word play in the use of proverbs in written discourse. A set of 303 occurrences of six English proverbs was collected in the Corpus of Contemporary American English and the non-canonical occurrences were analysed and classified. It appears that most of these manipulations are simple contextual adaptations including noun-phrase substitutions and only very few occurrences could qualify as instances of wordplay. To verify this, a questionnaire with 32 of the non-canonical occurrences was administered to a group of 12 native speakers who rated them for humour and cleverness. A comparison of the five occurrences with the highest ratings and the five with the lowest ones confirmed that the simple contextual adaptation of proverbs does not create wordplay which requires semantic complexity combined with humour.

A. Rassi, J. Baptista, O. Vale describe a methodology for identifying proverbs automatically and their variants in running texts. This methodology is based on existing compilations of proverbs, by exploring the regular syntactic structures that most proverbs present and intersecting syntactic structure with the lexical units of the proverbs. From the syntactic regularities we divided the data into 13 different classes. Finite-state automata is used to represent the regular patterns found in the classes. The results showed a precision rate of 74.68% tested in Brazilian Portuguese journalistic corpus.

A. Krikmann studies the corpora of Estonian dialect words, riddles and proverbs and reveals two well-defined areas of language and culture and one less salient:

- South-Eastern Estonia in abroad sense (Se+Vo+TaL) together with the less concentrated Mulgi region
- The West-Estonian Islands together with the less concentrated Western and North-Western Estonia
- The Northern and North-Eastern Coast of Estonia

The study is introduced by a brief overview of the frequency distributions of cultural versus geographic units, relationships between λ-coefficients, correlation coefficients and euclidean distances and the distances between geographic units as well as of the geographical distribution of rare versus common material (Krikmann, 2014).

RESULTS AND DISCUSSION

Jesensek (2013a) investigates the lexicographic example within the scope of paremiography (proverb lexicography). Inter disciplinarily and through inclusion of the phraseological and paremiographical theoretical knowledge of semantics, pragmatics and grammar of the proverbs, assertions are then developed on the quality characteristics of text passages with the help of which potential lexicographic examples within the scope of paremiography can be identified, systematically evaluated and selected. Finally the acceptability and operationability of the determined quality characteristics are discussed as well as some further research questions addressed. The considerations are based on the experiences from the development of a multilinguistic pareomiographical product that was conceptualized and developed as documentation of the actual proverbial use and also as learning and teaching material in foreign language learning contexts. The article will therefore contribute to the development of a theory of the lexicographic example and until now not yet realized within the scope of paremiography (Jesensek, 2013a). His another research relates to proverbs in the contemporary language use. Empirical corpus-linguistic data attest a significantly high incidence of proverbs in many communicative

domains. Were they once considered primarily a stylistic, rhetorical and didactic device, the present use of proverbs shows a clear formally-structural as well as functional change. This change is evident in a frequent innovative and creative as well as playful textual insertion that is largely established by their diverse variational and transformational potential and has usually deliberate stylistic, pragmatic and functional influences as a consequence. From the perspective of (foreign) language didactics, the phenomenon should be of particular interest and integrated into language learning. However, this is largely not yet the case. It is perhaps that old traditionalistic attitudes towards proverbs and presumptive lack of knowledge about the formally-structural as well as semantically-pragmatic features of proverbs play a role. The main objective of this study is to point to those features of proverbs that are of relevance for the (foreign) language didactics and should have consequences for language learning (Jesensek, 2013b).

Al-Momani and Jaradat (2012) examine Jordanian proverbs to show by examining a corpus that is well-defined, how these proverbs construe the other in all its difference. The analysis of the material reveals that the other has several possible identities in Jordanian Proverbs and that these identities are represented by various stereotypical images. It also demonstrates that the other falls under multiple categorizations which are expressed by diverse linguistic strategies.

Rozumko (2012) investigates recent borrowings of English proverbs into Polish. The present study is corpus based. It discusses the contexts in which the proverbs appear in Polish, the metalinguistic tags used to introduce them and the cultural significance of these borrowings.

Estaji and Nakhavali (2011) study corpora of animal expressions in English and Persian. In this study, “dog” expressions are examined based on Hsieh’s approach of semantic molecules to explore the salient meanings and the cultural backgrounds. Animal expressions may reveal people’s thoughts, emotions, culture and customs. The analysis of about 10,000 Persian and English proverbs shows that there are 207 Persian and 97 English “dog” expressions. In spite of cultural and social differences between English and Persian, the salient semantic properties derived from the name of this animal are nearly the same. The main semantic molecules of the word “dog” are “worthless, bad-tempered, cruel, violent” in both English and Persian.

On the basis of French, Hungarian, English, German and Russian corpora of anti-proverbs Barta *et al.* (2009) examine word play based on polysemy, homonymy and homophony. Then, they explore the use of proper nouns in proverb transformations based on polysemy, homonymy and homophony.

CONCLUSION

The given research provides analyses of approaches to proverbs and sayings study of different languages and overviews how proverbs and sayings research can benefit from corpus-based approach. The most essential advantage of applying corpora in linguistic researches is its reliable empirical basis, constant updating, and availability of proof of using a lexical unit in different meanings from a multiple number of sources of different genres.The corpus linguistic perspective also shows that proverbs themselves can be realizations of more general patterns and schemas what makes corpus-based researches valuable for studies in such diverse fields as pragmatics, semantics, grammar, stylistics, syntax, lexicography, paremiography, cultural studies. Work with large data basis enables the possibility of revealing regularities in many similar cases of usage and eliciting some specific structures which were not so obvious before work with corpus. Corpus-based studies are beneficial for researches from the point of view of finding unusual cross-connections and unexpected relations, which are also a great contribution to the larger picture of language and vocabulary.

ACKNOWLEDGEMENT

The research is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

Al-Momani, R. and R. Jaradat, 2012. Representations of the other in jordanian proverbs. Dirasat: Hum. Soc. Sci., 39: 826-834.

Barta, P., H. Hrisztova-Gotthardt, A. Litovkina and V.K. Polysemy, 2009. Homonymy and homophony in anti-proverbs-with French, Hungarian, English, German and Russian examples. Acta Ethnographica Hungarica, 54: 63-74.

Durco, P., 2014. Paremiology and corpus linguistics. Bulletin of Novgorod State University. http://cyberleninka.ru/article/n/paremiologiya-i-korpusnaya-lingvistika.

Estaji, A. and F. Nakhavali, 2011. Contrastive analysis of dog expressions in English and Persian. Us-China Foreign Language, 9: 213-219.

Jesensek, V., 2013a. Use of proverbs today. Linguistics and language didactics considerations. Muttersprache, 123: 81-98.

Jesensek, V., 2013b. The lexicographic example in paremiography. Forms and functions. Lexikos, 23: 150-171.

Krikmann, A., 2014. On the areal division of Estonia according to dialect and folklore material. Keel Ja Kirjandus, 8-9: 708-721.

Krotova, E.O., 2016. A corpus-based approach to semantics of German idioms. http://youngresearchers journal.org/wp-content/uploads/2013/12/Krotova-E-semantika-idiom.pdf.

Matej, M., 2015. The units of Slovene paremiological minimum in the corpus of spoken Slovene (GOS). Slavisticna Revija, 63: 1-15.

McEnery, T., R. Xiao and Y. Tono, 2006. Corpus-based Language Studies: An Advanced Resource Book. Taylor and Francis, USA., ISBN: 9780415286237, Pages: 386.

Rozumko, A., 2012. English influence on Polish proverbial Language. In: The Anglicization of European Lexis, Furiassi, C., V. Pulcini and F.R. Gonzalez (Eds.). John Benjamins Publishing, USA., pp: 261-277.

Zavyalova, N.A., 2013. Idioms and politics: Old age antiquity or burning isssues of present? Political Linguistics. http://cyberleninka.ru/article/n/idiomy-i-politika-sedaya-starina-ili-ostreyshie-voprosy-povsednevnosti.

Zirker, A. and E. Winter-Froemel, 2015. Wordplay and Metalinguistic/Metadiscursive Reflection: Authors, Contexts, Techniques and Meta-Reflection. De Gruyter, USA., pp: 135-159.