DARIA DAYTER
https://orcid.org/0000-0001-5054-5330
Tampere University

ŁUKASZ GRABOWSKI
https://orcid.org/0000-0002-3968-9218
Uniwersytet Opolski

CORPUS-BASED AND CORPUS-DRIVEN RESEARCH ON TRANSLATION AND INTERPRETING IN RUSSIAN: THE PAST, THE PRESENT AND THE FUTURE

Recently, there has been a growing interest in descriptive and applied studies on translation conducted with the use of methodologies grounded in corpus linguistics. In this paper, we present an overview of state-of-the-art research in corpus-based and corpus-driven translation and interpreting studies conducted with the use of Russian corpora, notably parallel and comparable ones. In contrast to research conducted on other European languages (English, French, German, Spanish etc.), the considerable research tradition in Russian remains relatively unknown to a wider scholarly audience. We outline the scope of research conducted so far, present the most important parallel and comparable corpora with a Russian subcorpus used in translation and interpreting studies, discuss the state-of-the-art research methods, including descriptive and inferential statistics, and summarize selected studies of considerable impact in the discipline. Finally, we identify research gaps and outline avenues for future research.

1. INTRODUCTION

The claim put forward almost thirty years ago by Albrecht Neubert and Gregory M. Shreve[1] that contemporary translation studies represent a cluster of overlapping disciplines and that there is no single way or approach to study translational phenomena is even more pertinent nowadays. One of the research strands that have gained mo-

---

[1] A. Neubert, G. Shreve, *Translation as Text*, The Kent State University Press, Kent-London 1992, p. 6.

mentum since the 1990s and early 2000s, i.e. since the publication of classic works by James Holmes[2], Mona Baker[3], Miriam Shlesinger[4], Maria Tymoczko[5], Sara Laviosa[6] and Maeve Olohan[7], among others, is descriptive studies on translation and interpreting conducted with the use of corpus linguistic methodologies (corpus-based and corpus-driven ones). This research direction offers new ways of conceptualizing and studying translation and interpreting largely from, but not limited to, a quantitative, statistical perspective,[8] also in an attempt to introduce more methodological rigour and objectivity. According to Olohan[9], corpus linguistic research methods allowed, among others, to reveal what is probable and typical in translation and combine quantitative and qualitative research methods in linguistic descriptions encompassing lexical, syntactic and discoursal features of translational and non-translational texts.

Three main research orientations in corpus linguistic descriptive translation studies have become particularly popular in the last three decades or so. These are research on translation/interpreting universals, translator's/interpreter's style and translation/interpretation style.[10] As for translation universals, they constitute generalizations based on repeatedly observed textual characteristics (lexical, gram-

---

[2] J. Holmes, *The Name and Nature of Translation Studies*, in: Lawrence Venuti (Ed.), *The Translation Studies Reader*, Routledge, London 1988/2000, pp. 172–185.

[3] M. Baker, *Corpus Linguistic and Translation Studies: Implications and applications*, In: M. Baker, G. Francis, and E. Tognini-Bonelli (Eds), *Text and Technology*, *in Honour of John Sinclair*. John Benjamins, Amsterdam 1993, pp. 233–250; M. Baker, *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*, "Target" 1995, no. 7(2), pp. 223–243.

[4] M. Shlesinger, *Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies*, "Meta" 1998, no. 43(4), pp. 86–493.

[5] M. Tymoczko, *Computerized Corpora and the Future of Translation Studies*, "Translators' Journal" 1998, 43(4), pp. 652–660.

[6] S. Laviosa, *Corpus-based translation studies: theory, findings, applications*. Rodopi, Amsterdam 2002.

[7] M. Olohan, *Introducing Corpora in Translation Studies*. Routledge, London 2004.

[8] This reflects the claim made by Holmes (1988/2000) who argues that descriptive Translation Studies should aim to describe, predict and explain translational phenomena using authentic empirical data with a goal of producing a systematic description of a translation product (text), process and function.

[9] M. Olohan, *Introducing Corpora in Translation Studies...*

[10] These research orientations mentioned in this paragraph are discussed in greater detail in Grabowski (2015).

matical, discoursal etc.) typical of translated texts irrespective of particular pairs of source and target languages[11], while the translator's style represents a personal attribute conceptualized as an individual and consistently used set of features of translation (or interpreting) that it shares with other translations (or interpretations) completed by one and the same translator or interpreter, irrespective of source texts.[12] Finally, translation style has been seen as a textual attribute rather than a personal one and it refers to the description of the way a translator or interpreter responded to a source text.[13] In the last two approaches, the style is largely conceptualized as "a property of texts constituted by an ensemble of formal features which can be observed quantitatively and qualitatively" and its analysis may involve "any linguistic feature that can be formally defined and measured computationally."[14]

Recently, corpus-based translation and interpreting studies have enjoyed considerable and growing popularity.[15] A perfunctory inspection of the most recent studies in the field (Kotze[16], Dayter[17]; Delaere, and De Sutter[18]; De Sutter, and Lefer[19]; Kajzer-Wietrzny,

[11] M. Baker, *Corpus Linguistic and Translation Studies…*

[12] J. Rybicki, *Stylometric translator attribution: do translators leave lexical traces?*, T. Piotrowski, Ł. Grabowski (Eds), *The Translator and the Computer*. Wydawnictwo Wyższej Szkoły Filologicznej, Wrocław 2013, pp. 193–204.

[13] K. Malmkjær, *What Happened to God and the Angels: An Exercise in Translational Stylistics*. "Target" 2003, no. 15(1), pp. 37–58; G. Saldanha, *Translator style. Methodological considerations*, "The Translator" 2011, 17(1), pp. 25–50.

[14] J.B. Herrmann, K. van Dalen-Oskam, and Ch. Schöch, *Revisiting Style, a Key Concept in Literary Studies*, "Journal of Literary Theory" 2015, pp. 43–44.

[15] M.-A. Lefer, *Parallel Corpora*. In: M. Paquot, and S. Gries (Eds), *A Practical Handbook of Corpus Linguistics*, Springer Nature Switzerland AG, Cham 2020, pp. 258.

[16] H. Kotze, *Converging what and how to find out why: An outlook on empirical translation studies*, L. Vandevoorde, J. Daems, and B. Defrancq (Eds), *New Empirical Perspectives on Translation and Interpreting*. Routledge, Abingdon 2020, pp. 333–370.

[17] D. Dayter, *Describing lexical patterns…*; D. Dayter, *Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event*. "Meta: Translator's Journal" 2021, no. 66(3), pp. 594–617; D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English: a corpus study*, In: L. Vandevoorde, J. Daems, and B. Defrancq (Eds), *New Empirical Perspectives on Translation and Interpreting*, Routledge, London 2020, pp. 67–91.

[18] I. Delaere, and G. De Sutter, *Applying a Multidimensional, Register-Sensitive Approach to Visualize Normalization in Translated and Non-Translated Dutch*, "Belgian Journal of Linguistics" 2013, no. 25, pp. 43–60.

[19] G. De Sutter, and M.-A. Lefer, *On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach*, "Perspectives" 2020, no. 28(1), pp. 1–23.

and Grabowski[20]) attests to the growing awareness among researchers, notably those specializing in variationist linguistics, of the limitations of basic quantitative (descriptive statistics) and qualitative (concordance analysis) research methods. These have turned out to be insufficient to address more complex problems of linguistic variation in translation and interpreting, and researchers instead resort to more sophisticated statistical methods (multivariate statistics etc.). This point has been elaborated recently by Stefan Th Gries[21], Gert DeSutter and Marie-Aude Lefer[22], and Maria Calzada Pérez and Sara Laviosa[23]: the latter two argue that monofactorial (i.e. based on a single independent explanatory variable) descriptive studies conducted by corpus linguists, e.g. related to frequency differences of a certain linguistic feature (dependent variable) between translational and non-translational texts (independent explanatory variable), do not contribute much to our understanding of translation or interpreting. In fact, there are a number of other factors that impact the use of language in a particular way in translation or interpreting and that is why it is necessary to use more sophisticated statistical methods[24] (e.g. regression models) to account for multifactorial empirical phenomena[25]. Thus, identification and comparisons of frequencies of individual linguistic features or pairs of source-text items and their translation equivalents in texts or corpora do not offer a plausible explanation of the observed phenomenon. In other words, putting

---

[20] M. Kajzer-Wietrzny, and Ł. Grabowski, *Formulaicity in constrained communication: An intermodal approach*, "MonTI" 2021, no. 13, pp. 148–183.

[21] S. Gries, *The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models*, *Corpora* 2015, no. 10(1), pp. 95–125; S. Gries, *On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally*, "Journal of Second Language Studies" 2018, no. 1(2), pp. 276–308.

[22] G. De Sutter, and M.-A. Lefer, *On the need for a new research agenda...*

[23] M. Calzada Pérez, and S. Laviosa, *Twenty-five years on: Time to pause for a new agenda for CTIS*. In: M. Calzada Pérez, and S. Laviosa (Eds), *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning*: *A critical reflection*, "MonTI" 2021, no. 13, p. 20.

[24] For more on statistical modeling of linguistic data (linear models, generalized linear models and mixed models etc.), see Winter (2019).

[25] This constitutes an interesting counterpoint to the developments in corpus linguistics, where there have been recent calls for backgrounding sophisticated statistical reporting in favour of more traditional linguistic analysis (Egbert, Larsson, and Biber 2020; Larsson, Egbert, Biber forthcoming). Also, a critical reflection on quantitative approaches to linguistic research, replication crisis etc. can be found in Sönning and Werner (2019).

forward an explanation without the use of multifactorial statistics is tantamount to mere speculation and should be discouraged in corpus linguistic research on translation and/or interpreting.

In what follows, we will present and discuss the scope of state-of-the-art corpus linguistic research on translation and interpreting conducted with the use of Russian language material (and set it against the backdrop of the recent advances in the field). We will also present the most important parallel and comparable corpora[26] of Russian compiled so far and used by researchers interested in descriptive translation and interpreting studies. We believe that while the work in the main European scholarly languages such as French and German has been recognised and well-integrated in the research landscape of translation and interpreting studies, e.g. Marianne Lederer[27] or Silvia Kalina[28], the considerable research tradition in Russian remains largely unknown.

2. CORPUS-BASED TRANSLATION STUDIES: AN OVERVIEW OF CORPORA WITH RUSSIAN LANGUAGE

Russian linguistics has developed from a strong philological tradition that comprises linguistic and literary studies under one umbrella, often locating them in the same research department. As a result, linguists have often worked on literary data to make their research relevant and applicable in this broader disciplinary context. This situation has led to proliferation of parallel corpora of literary language, and their relative prominence compared to other European languages' resources[29].

Among freely available resources, Russian corpus studies boast a Russian-French parallel corpus of poetry[30] whose web interface enables both the traditional queries (lemma, wordform) and searching

---

[26] See Zanettin (2012), Mikhailov and Cooper (2016), Ustaszewski and Stauder (2020), among others, for a more detailed discussion of the typology and compilation of various types of corpora used in translation studies research.

[27] M. Lederer, *La traduction simultanée. Expérience et thé*orie, Lettres modernes Minard, Paris 1981.

[28] S. Kalina, *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*, Günter Narr, Tübingen 1998.

[29] For a more detailed overview of Russian corpora, including monolingual and comparable ones, see Kopotev et al. (2021).

[30] A. Balakin, A. Bodrova, S. Orekhov, and M. Rybina. Russian-French Poetry Corpus 1800–1820, 2011, Available at: http://nevmenandr.net/fr/.

specifically in the rhyming area. Another freely available parallel corpus of fiction is the Chinese and Russian Literature Corpus (http://www.rucorpus.cn/), containing a subcorpus of Anton Chekhov's short stories (Ru>Zh) and a subcorpus of Chinese fiction (Zh>Ru). The literary component of the Polish-Russian Parallel Corpus[31] incorporates both bidirectional translations Ru<>Pl and translations of literary works from other languages into Russian and Polish. An unusual resource at the intersection of historical and comparative linguistics is the parallel corpus of translations of *The Tale of Igor's Campaign* (*Слово о полку Игореве*), an epic poem in Old Slavic.[32] It includes around a hundred intralingual translations into Modern Russian as well as 44 translations into other languages.

All four corpora lend themselves well to interdisciplinary use because they offer intuitive web interfaces. A great number of further parallel corpora of fiction, often compiled for specific research projects, are freely available, but require the researcher to be somewhat technically savvy: they offer a download option rather than a web interface. Among these we could highlight the PELCRA Polish-Russian parallel corpus with 20 Polish and 14 Russian literary works, with 4.25 million words[33]; the Chinese-Russian corpora of individual novels, e.g. the hand-aligned corpus for *The Republic of Wine*[34] (Piao et al. 2014); HunOr: A Hungarian–Russian Parallel Corpus[35], with sam-

---

[31] M. Łaziński, M. Kuratczyk, B. Orekhov, and E. Słobodjan, *The Polish-Russian Parallel Corpus and Its Application in the Linguistic Analysis*, "Prace Filologiczne" 2012, vol. LXIII, pp. 209–218; M. Łaziński, and M. Kuratczyk, *Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego*, E. Gruszczyńska, and A. Leńko-Szymańska (Eds), *Polskojęzyczne korpusy równoległe*. Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2016: 83–96; M. Łaziński, and M. Kuratczyk, *Польско-русский параллельный корпус Варшавского университета*, "Język i Metoda», 2016, no. 3, pp. 85–93].

[32] Б. Орехов, *Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы*, В. Плунгян (Ed.). *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы*. Санкт-Петербург: Нестор-История, 2009, pp. 462–473.

[33] P. Pęzik, M. Ogrodniczuk, and A. Przepiórkowski, *Parallel and spoken corpora in an open repository of Polish language resources*, Human Language Technologies as a Challenge for Computer Science and Linguistics. LTC, Poznań 2011, http://pelcra.pl/new/polrus (30.08.2021).

[34] Z. Piao, Q. Li, and L. Wang, *Jiyu han-e pingxing yuliaoku de „jiu guo" dieyinci eyi guilü yanjiu*, "Zhongguo eyu jiaoxue" 2014, no. 3, pp. 46–51.

[35] M. Szabo, V. Vincze, and I. Nagy, *HunOr: A Hungarian–Russian Parallel Corpus*, In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA),

ples of literary novels, scientific texts and official documents (with approx. 800,000 tokens in each language); or the Polish-Bulgarian-Russian Parallel Corpus[36] including 55 texts (literary novels and legal treaties) translated into the aforementioned languages.

Another type of parallel corpora presents the register of 'language for special purposes', e.g. business texts, academic language, political and diplomatic discourse, news articles or marketing texts. These can be found in the PEST corpus with state treaties in Swedish, Russian, Finnish[37], in the Corpus of Russian Translations of Social and Political Works, including translations of English and German 18th-century treaties into Russian[38], as well as in the UMC 0.1 Czech-Russian-English Multilingual Corpus with internet texts and news articles[39]. One can also find specialized parallel texts with Russian language data in the constantly growing Open Parallel Corpus collection[40] that enables one to browse and download parallel text files in various formats (e.g. txt, tmx). Another genre corpus is the Russian Learner Translator Corpus (RuLTC), which comprises translations of newspapers and other English mass-media performed by advanced students of translation. RuLTC contains approx. 2.3 million tokens and has an online search interface freely available to researchers.[41]

Finally, a number of large, multi-register corpora are available to the researchers that comprise different text types: fiction, lan-

Istanbul 2012, pp. 2453–2458, http://www.lrec-conf.org/proceedings/lrec2012/pdf/262_Paper.pdf.

[36] A. Kisiel, V. Koseska-Toszewa, N. Kotsyba, J. Satoła-Staśkowiak, and W. Sosnowski, *Polish-Bulgarian-Russian Parallel Corpus*, CLARIN-PL digital repository, 2016, http://hdl.handle.net/11321/308.

[37] M. Mikhailov, M. Santalahti, and J. Souma (Eds.), *PEST. A parallel electronic corpus of state treaties*, John Benjamins, Amsterdam 2019.

[38] Polskoy, Sergey, and Rjéoutski, Vladislav, *The Corpus of Russian Translations*, 2021. https://krp.dhi-moskau.org/ru/translations (27.08.2021).

[39] N. Kluyeva, and O. Bojar, *UMC 0.1: Czech-Russian-English Multilingual Corpus*. Proceedings of International Conference Corpus Linguistics, October 2008, pp. 188–195, https://ufal.mff.cuni.cz/umc/cer/download.php?f=umc-0.1-paper-2008.pdf.

[40] J. Tiedemann, *OPUS - Parallel Corpora for Everyone*, "Baltic Journal of Modern Computing" 2016, no. 4 (2): Special Issue: *Proceedings of the 19th Annual Conference of the European Association of Machine Translation (EAMT)*..

[41] A. Kutuzov, and M. Kunilovskaya. *Russian Learner Translator Corpus*, in: P. Sojka, A. Horák, I. Kopeček, K. Pala (Eds.), *Text, Speech and Dialogue*. TSD 2014. Lecture Notes in Computer Science, vol. 8655. Springer, Cham, https://doi.org/10.1007/978-3-319-10816-2_39.

guage for specific purposes, and sometimes even transcriptions of spoken language. These will be well familiar to corpus linguists. There is the Russian National Corpus[42] (https://ruscorpora.ru/, NKRY 2005) of 300 million words covering a breadth of historical periods and genres and offering a large POS-tagged component (manually checked); and ParaSol,[43] a project that has similar aims to the Russian National Corpus but is currently at an earlier development stage, and focussing specifically on Slavic languages. Another prominent example of this family is InterCorp 13,[44] an enormous parallel corpus of 1.551 million words with Czech as a pivot language and 40 further languages, including Russian and many more Slavic languages.

3. CORPUS-BASED INTERPRETING STUDIES: AN OVERVIEW OF RUSSIAN LANGUAGE CORPORA

Corpus-based approaches to interpreting studies have become the new standard, compared to the earlier introspection-based approaches. However, as late as in 2011, Setton still called Corpus-Based Interpreting Studies (CIS) "a cottage industry." This characteristic referred to the fact that CIS was represented by many disparate studies, comparatively small-scale, rather than a single concerted effort with uniform norms, transcription conventions, and systematic research agenda.

Interestingly, corpus-based interpreting research in Russian goes back several decades. For example, Ghelly Chernov's groundbreaking research on anticipation in simultaneous interpreting is based on corpus data.[45] According to the compilers of Chernov's

---

[42] As such, NKRY contains a number of parallel sub-corpora that have been developed since 2005 (Sitchinava 2012).

[43] R. von Waldenfels, *Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment*, in: B. Brehmer, V. Zdanova, R. Zimny (Eds), *Beiträge der Europäischen Slavistischen Linguistik*: POLYSLAV 9. Verlag Otto Sagner, München 2006, pp. 123–138.

[44] F. Čermák, and A. Rosen, *The case of InterCorp, a multilingual parallel corpus*. "International Journal of Corpus Linguistics" 2012, no. 13(3), pp. 411–427.

[45] Gh. Chernov, *Inference and Anticipation in Simultaneous Interpreting*, Benjamins, Amsterdam 1978/2004.

recently translated edition Setton, and Hild,[46] Chernov worked with a "shoebox corpus" (a paper-based card index) of the United Nations speeches which he taped and transcribed himself. The tapes were allegedly made available to the publishing house in 1987, but it appears that now both the tapes and the unpublished transcriptions are lost[47]. Chernov's colleague and co-author Irina Zimnaya also worked with corpora of simultaneous interpreting[48], although we have no record of this resource. Since much of the publishing effort at the time was pedagogically oriented, we have the resulting textbooks based on the corpus research[49] but not the corresponding corpora.

Although CIS has been gaining momentum in the last decade, at the moment of writing, there are still very limited corpus resources available for the Russian language. This scarcity can be explained by the fact that most existing interpreting corpora are based on the EUROPARL database, a large parallel corpus extracted from the proceedings of the European Parliament and including 21 European languages. Since the Russian Federation is not a part of the European Union and is therefore not represented in the European Parliament, the resource does not contain a Russian component.

A somewhat similar source of parallel language data including the Russian language is provided by the records of the United Nations' events and proceedings. Although transcripts of the UN proceedings are not readily available, researchers have access to United Nations Web TV (http://webtv.un.org/). The website airs a variety of UN events such as General Assembly, press conferences, press briefings etc., with simultaneous interpretation into the official UN languages in a separate audio channel. The events are aired live and the recordings, including simultaneous interpretation, are stored on the website. The material is in public domain, and the United Nations Publications allows the use of portions of its content for educational and research purposes. While the material needs to

---

[46] R. Setton,, and A. Hild, *'Editors' critical foreword,* in: Gh. Chernov (Ed.), *Inference and Anticipation in Simultaneous Interpreting*, Benjamins, Amsterdam 2004, pp. IX–XXII.

[47] Ibidem.

[48] I. Zimnaya, and Gh. Chernov, *Verojatnostnoe Prognozirovanie v Processe Sinkhronnogo Perevoda.* In: A. Leont'ev, N. Zhinkin, and A. Shahnarovich (Eds.), *Predvaritel'nye materialy jeksperimental'nyh issledovanij po psiholingvistike*, IY SSSR, Moscow 1973.

[49] Eg. A. Shiryaev, *Sinkhronnyj perevod*, Voenizdat, Moscow 1979.

be laboriously transcribed from the video/audio format, this additional step results in cleaner data than the automatically harvested EUROPARL texts.[50]

One existing corpus that includes a Russian component and is based in part on the UN WebTV data is the SIREN.[51] The SIREN is a bidirectional parallel Russian-English corpus of original texts and their simultaneous interpreting into English with circa 235,000 word tokens, transcribed, annotated, POS-tagged and parsed (at the sentence-clause level). The SIREN includes 21 speech events from the WebTV archives, among them, for example, the UN Security Council's meetings or sessions of Universal Periodic Review.[52] Although the corpus is not freely available in a language depository at the moment of writing, the creator can be contacted to provide the corpus on request.

While the SIREN represents the genre of political discourse, another resource, CoInCoUT (Court Interpreting Corpus at the University of Tampere), focuses on courtroom discourse. CoInCoUT includes Finnish-Russian interpretation of 9 sessions, approx. 49,000 words, with three different interpreters.[53]

Other existing interpreting corpora with a Russian component are not available to researchers, but are purpose-built for specific studies and described in unpublished PhD dissertations.[54]

---

[50] Corpora created from the transcripts of video recordings and the parallel interpreting audio stream are not to be confused with the parallel corpora of the UN written documents, e.g. The United Nations Parallel Corpus 1.0 (Ziemski et al. 2016).

[51] D. Dayter, *Describing lexical patterns...*; D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English: a corpus study*, in: L. Vandevoorde, J. Daems, and B. Defrancq (Eds.), *New Empirical Perspectives on Translation and Interpreting*, Routledge, London 2020, pp. 67–91.

[52] D. Dayter, *Collocations in non-interpreted and simultaneously interpreted English...*

[53] M. Mikhailov, H. Tommola, and N. Isolahti, *Spoken to Spoken vs. Spoken to Written: Corpus Approach to Exploring Interpreting and Subtitling*, "Polibits" 2010, no. 41, pp. 5–9.

[54] E.g. A. Gorshkova, *Sopostavitel'noye issledovaniye sposobov dostizheniya ekvivalentnosti v sinkhronnom i pis'mennom perevodakh*. Unpublished Phd dissertation, Rossiyskiy universitet druzhby narodov, Moskva 2004; V. Ilukhin, *Strategii v sinkhronnom perevode*. Unpublished PhD dissertation, Moscow 2001; I. Gurin, *Priyemy rechevoy kompressii prisinkhronnom perevozhde s russkogo yazyka na angliyskiy*. Unpublished Phd dissertation, Moskovskiy gosudarstvennyy universitet, Moskva 2009.

To sum up, while translation corpora are a well-developed segment in Russian corpus linguistics resources, there is still a distinct lack of freely available corpora of interpreting.

4. RESEARCH ON TRANSLATION AND INTERPRETING: THE PAST AND THE PRESENT

A perfunctory inspection of state-of-the-art corpus research on translation and interpreting in Russian shows that the most prominent studies can be thematically grouped into two main areas, namely the studies dealing with identification of specific features of translation and/or interpreting and the studies comparing different translation/ interpreting modes.

The first kind of studies typically draw their inspiration from the theories of translation universals[55] or from the translation/interpreting 'strategies'.[56] Both of these attempt to capture universal patterns of linguistic behaviour in translation, which manifest themselves on the level of lexis, syntax, and discourse. For example, early cognitive research focused on the strategy of anticipation in simultaneous interpreting.[57] Relying on a pre-computer index-card corpus, the researchers described multiple instances of anticipation in French-Russian interpreting and concluded that strategic anticipation is a result of the message redundancy characteristic of all natural languages. One of the resources described in section 3 above, the SIREN corpus, gave rise to comprehensive corpus-based research on universals and strategies in Russian-English simultaneous interpreting. The findings are somewhat contradictory with regard to universals (although one of the language subcorpora conforms to the expectations of translation theory, the other exhibits the opposite trend) and invite further corpus-based investigation of interpreting[58].

---

[55] M. Baker, *Corpora in Translation Studies: An Overview and Some Suggestions for Future Research*, "Target" 1995, no. 7(2), pp. 223–243; A. Chesterman, *Hypothesis about translation universals*, in: G. Hansen, K. Malmkjær, and D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies*, John Benjamins, Amsterdam 2004, pp. 1–13.

[56] S. Kalina, *Strategische Prozesse beim Dolmetschen…*

[57] I. Zimnaya, and Gh. Chernov, *Verojatnostnoe Prognozirovanie v Processe Sinkhronnogo Perevoda…*

[58] D. Dayter, *Describing lexical patterns in simultaneously interpreted discourse…*; D. Dayter, *Strategies in a corpus of simultaneous interpreting…*

The descriptions of linguistic markers that serve to operationalise universals and strategies have been applied in the second thematic area of translation corpus research, the studies comparing various translation and interpreting modes. For example, the work by Lapshinova-Koltunski and the RuLTC team look at translation universals (translationese) that manifest to different degrees e.g. in human vs. machine translation, or performed by novice vs. experienced translators.[59]

A comparison between novice (learner) and professional Russian translations of English mass-media texts with the reference Russian corpus of non-translations (i.e. native texts of the same genre collected in the NKRY), undertaken in order to identify lexical differences between the three text varieties, was the focus of corpus research by Kunilovskaya et al.[60] The study was grounded in the concept of translation universals. Using selected quantitative indicators of style (sentence length, frequencies of morphological forms of word classes, lexical variety measured by the type/token ratio and the proportion of high-frequency words, and lexical density measured by the ratio of content words to the total number of words in texts) and suitable tests of statistical significance and effect size metrics[61] found that, first, novice translations, extracted from the aforementioned Russian Learner Translator Corpus,[62] differ more from native texts than translations performed by professionals (published by 10 selected Russian news portals), and, second, that both novice and professional translations revealed characteristics that made them linguistically different from native, non-translational texts (in terms of proportion of high frequency words, lexical density and specific frequency of

[59] M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian: Lexical profiles*, "Translation, and Interpreting" 2018, no. 10(1), pp. 33–52, https://trans-int.org/index.php/transint/article/view/585/304 (30.08.2018); M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality in English-Russian Human Translation*, in: I. Temnikova, C. Orasan, G. Corpas Pastor,, and R. Mitkov (Eds.), *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology* (HiT-IT 2019), Varna, Bulgaria, September 5–6 2019, pp. 47–56, https://www.aclweb.org/anthology/W19–8706.pdf (30.08.20210.

[60] M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian...*; M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality...*

[61] M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian: Lexical profiles...*

[62] A. Kutuzov, and M. Kunilovskaya, *Russian Learner Translator Corpus...*

word forms).[63] In future research, it would be worthwhile verifying whether the same tendencies are observed in language pairs other than English-to-Russian, for example in the Polish-to-Russian translation conducted between more typologically similar Slavic languages. This and other avenues for future research will be discussed in the following Conclusions section.

The findings of this research are complemented by Kunilovskaya, and Lapshinova-Koltunski,[64] a study which compares originally produced vs. translated texts and good vs. bad students translations in terms of a theoretically motivated set of features. This set of features is based on the literature in the field of register studies that are known to capture translationese, supplemented by a number of morphosyntactic features that are meant to reflect fluency. The proposed features are a combination of relatively shallow lexico-grammatical indicators and of features that could be automatically tagged by a Universal Dependencies tagger, 45 in all. For instance, the authors included seven Universal Dependency features that have been shown to be translationese indicators specifically for the Russian-English pair: adjectival clause, auxiliary, passive voice auxiliary, clausal complement, subject of a passive transformation, asyndeton, a predicative or clausal complement without its own subject. In addition to the student translations from the RuLTC corpus and the professional translations from a range of established electronic media, the authors used original Russian newspaper texts as a reference corpus. The student translations were additionally evaluated for quality by translation experts, with only the top three best ranking and the bottom three worst rankings used in the analysis. For the study, the authors used a supervised learning model to solve two classification tasks based on the extracted feature frequencies: (1) translated vs. non-translated texts, and (2) best-ranking vs. worst-ranking translations. Overall, Kunilovskaya, and Lapshinova-Koltunski[65] found that the chosen features can distinguish between translated and non-translated texts with near perfect accuracy, that is, that these features reflect translationese reliably. The quality task, on the other hand, was not solved suc-

---

[63] M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian…*

[64] M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality…*

[65] Ibidem.

cessfully: the algorithm performed barely above the chance level. The authors conclude that translationese does not equal a reduction in quality; this might be due to the fact that 'quality' is more than lexico-grammatical faithfulness to standard; it reflects semantic faithfulness and pragmatic acceptability as well.[66]

While the two studies we have just described in detail concern the newspaper register, other research demonstrates feasibility of corpus investigations into translationese in Russian in other genres. For example, Krasnopeeva[67], cited by Kunilovskaya et al.[68], looked at originally authored and translated Russian fiction and found statistically significant differences in lexis between the two. It is a promising avenue of research to (a) continue looking at the lexical, morphological and syntactic features distinguishing translated and non-translated Russian on the basis of large corpora; (b) continue looking into morphological, syntactic and textual features that impact register variation in translation;[69] and (c) to conduct corpus-based variationist research into "constrained language"[70] taking into account variables other than translation status. Such variables include, for example, translator's proficiency (investigated in the two studies above), but also the setting for the translation/interpreting process, the communication medium, or the translation's target audience.

5. CONCLUSIONS

In this paper, we focused on recent advances in corpus linguistic research on translation and interpreting, with particular emphasis on studies conducted using Russian language corpora. It goes without saying that there are other theoretical paradigms beyond empirical

---

[66] Ibidem.

[67] E. Krasnopeeva, *Lexical features of Russian translated discourse. A corpus-based comparative study of contemporary narrative prose*. Unpublished PhD dissertation, Chelyabinsk State University 2015 (cited in Kunilovskaya et al. 2018, pp. 34).

[68] M. Kunilovskaya, N. Morgoun, and A. Pariy, *Learner vs. professional translations into Russian...*

[69] M. Kunilovskaya,, and G. Corpas Pastor, *Translationese and Register Variation in English-To-Russian Professional Translation*, in: Vincent Wang, Lily Lim, and Defeng Li (Eds.). *New Perspectives on Corpus Translation Studies. New Frontiers in Translation Studies*, Springer, Singapore 2021 (https://doi.org/10.1007/978–981-16-4918-9_6).

[70] I. Lanstyak,, and P. Heltai, *Universals in Language Contact and Translation*, "Across Languages and Cultures" 2012, 13(1), pp. 99–121.

corpus linguistic approaches that contribute to a better understanding of translation and interpreting. Our critical overview revealed first and foremost a dearth of freely available corpus resources in interpreting that include a Russian component - the problem that can be traced back to the practical difficulty of creating interpreting corpora. In addition, the EUROPARL compendium, which serves as a source for most existing interpreting corpora, is not helpful in the case of Russian.

Although Russian corpus linguistics is a fruitful discipline that has existed since the 1980s, it is possible to identify new directions and research gaps in the studies of translation and interpreting that may be pursued further in the near future. One of the areas that has enjoyed considerable popularity in recent years is research on so-called constrained communication, where language use is mediated by translation/interpreting, foreign language use or both,[71] which is an underexplored area of study in the context of Russian language. Also, given the growing popularity of research on amateur, post-edited and L2 translation[72] (e.g. into Russian as a foreign language for translators/interpreters) conducted on main European languages, we may note — at least to the knowledge of the Authors — the scarcity of such corpora of Russian, notably in the Polish-Russian language pair. Likewise, there are no Russian translation learner corpora with a Polish component, or Russian translation corpora comprising professional and amateur translations with a Polish component. As in much of academic research, translation/interpreting corpora including English are overrepresented among available resources: for the English-Russian language pair, there exist e.g. Kutuzov, and Kunilovskaya; Dayter;[73] Kunilovskaya, and Lapshinova-Koltunski.[74] Thus, research into translation taking into account the translator qualifications represents an important research gap for languages other than English. This gap could be addressed in the future provided that suit-

---

[71] E.g. I. Lanstyak,, and P. Heltai, *Universals in Language Contact…*; H. Kruger & B. Van Rooy, *Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English*, "English World-Wide" 2016, no. 37(1), pp. 26–57; M. Kajzer-Wietrzny, and Ł. Grabowski, *Formulaicity in constrained communication…*

[72] This point has been also emphasized by Lefer (2021).

[73] A. Kutuzov, and M. Kunilovskaya, *Russian Learner Translator Corpus…*; D. Dayter, *Describing lexical patterns…*

[74] M. Kunilovskaya, and E. Lapshinova-Koltunski, *Translationese Features as Indicators of Quality…*

able corpora, representing those underexplored translation varieties, are developed. This may also contribute to the advancement of applied research on translation teaching and translation quality assessment, among others, in the Polish-Russian language pair.

Finally, it may be expected that more research will be conducted in the future using a variety of supervised or unsupervised machine learning techniques to study translational and interpreting phenomena through identification of hidden patterns of language use in large data collections. For example, Ustaszewski[75] showed how to apply supervised machine learning techniques (more precisely, Support Vector Machines (SVM) used for data classification) in order to computationally distinguish between direct translations from indirect translations based on a set of 26 linguistic features (i.e. represented as multidimensional vectors of numerical features). It is also clear that the growing popularity of machine learning techniques in the translation industry (deep learning, neural machine translation etc.) will have implications not only for the methodologies used by researchers studying translational phenomena, but also for the future of translation education.[76] This and other considerations promise a rise in popularity of interdisciplinary research and synergies between linguistics, literary studies, information technology / natural language processing, psychology etc. Advanced statistical and machine learning techniques will be brought to bear on corpus data to cast more light on the specificity of translation and interpreting, from the perspective of the product (text), process and function. These are only beginning to be explored for Russian, for example, in the work of Kutuzov, and Kuzmenko,[77] who used neural embedding models to study diachronic change. Collaborations between corpus and computational linguists may lead to new exciting research in the classic areas of interest we had identified in this paper.

In the same vein, we expect that there will be many more multifactorial studies[78] conducted in the future in Polish-Russian or Russian-

---

[75] M. Ustaszewski (2021) also presents important methodological considerations.

[76] G. Massey, and M. Ehrensberger-Dow, *Machine learning: Implications for translator education*. "Lebende Sprachen" 2017, no. 62(2), pp. 300–312.

[77] A. Kutuzov, and E. Kuzmenko, *Two centuries in two thousand words. Neural embedding models in detecting diachronic lexical changes*, in: Mikhail Kopotev, Olga Lyashevskaya, and Alto Mustajoki (Eds.), *Quantitative Approaches to the Russian Language*, Routledge, London 2017.

[78] See Paquot, and Gries (2020), notably chapters collected in Part V, for a more detailed discussion of statistical techniques.

Polish translation/interpreting research. Multifactorial approaches would enable one to posit explanatory hypotheses as to the text-external factors that govern the use of particular linguistic features (including, for example, particular translation equivalents) of translated, non-translated texts (native) or constrained texts. Monofactorial approaches that have been mostly applied so far in the studies of translationese are largely limited to the analysis (i.e. identification and description) of frequency distributions of particular linguistic features and of a single independent explanatory variable, and which, as such, only enable descriptive hypotheses. This future research direction, however, depends on whether researchers first develop suitable Polish-Russian translational and/or interpreting corpora, rich in metadata[79] (e.g. direction of translation, mode of delivery, text variety etc.), which is a *sine qua non* condition for such studies to flourish.

REFERENCES

Baker, Mona. "Corpus Linguistic and Translation Studies: Implications and applications." *Text and Technology: in Honour of John Sinclair*. Baker, Mona, and Francis, Gill, and Tognini-Bonelli, Elena (Eds.). Amsterdam: John Benjamins, 1993: 233–250.

Baker, Mona. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 1995, no. 7(2): 223–243.

Balakin, Alexey, and Bodrova, Alina, and Orekhov, Sergey, and Rybina, Maria. Russian-French Poetry Corpus 1800–1820 < http://nevmenandr.net/fr/>.

Bernardini, Silvia, and Ferraresi, Adriano, and Miličević, Maja. "From EPIC to EP-TIC — Exploring simplification in interpreting and translation from an intermodal perspective." *Target* 2016, no. 28: 61–86.

Calzada Pérez, Maria, and Laviosa, Sara. "Twenty-five years on: Time to pause for a new agenda for CTIS." *Reflexión crítica en los estudios de traducción basados en corpus / CTS spring-cleaning: A critical reflection*. Calzada Pérez, Maria, and Laviosa, Sara (Eds.). MonTI 13, 2021: 7–32.

Čermák, František, and Rosen, Alexandr. "The case of InterCorp, a multilingual parallel corpus." *International Journal of Corpus Linguistics*, 2012, no. 13(3): 411–427.

Chernov, Ghelly. *Inference and Anticipation in Simultaneous Interpreting*. Amsterdam: Benjamins, 1978/2004.

Chesterman, Andrew. "Hypothesis about translation universals." Hansen, Gyde, and Malmkjær, Kirsten, and Gile, Daniel (Eds.). *Claims, Changes and Challenges in Translation Studies*. Amsterdam: John Benjamins, 2004: 1–13.

---

[79] The European Parliament Translation and Interpreting Corpus (EPTIC) (Bernardini et al. 2016) or TransBank corpus (Stauder, and Ustaszewski 2020) are examples of translational/interpreting corpora rich in metadata.

Dayter, Daria. "Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN)." *FORUM* 2018, no. 16(2): 241–264 <https://doi.org/10.1075/forum.17004.day>.

Dayter, Daria. "Collocations in non-interpreted and simultaneously interpreted English: a corpus study." *New Empirical Perspectives on Translation and Interpreting*. Vandevoorde, Lore, and Daems, Joke, and Defrancq, Bart (Eds). London: Routledge, 2020: 67–91.

Dayter, Daria. "Strategies in a corpus of simultaneous interpreting. Effects of directionality, phraseological complexity, and position in speech event." *Meta: Translator's Journal* 2021, no. 66(3): 594–617.

Delaere, Isabel, and De Sutter, Gert. "Applying a Multidimensional, Register-Sensitive Approach to Visualize Normalization in Translated and Non-Translated Dutch." *Belgian Journal of Linguistics* 2013, no. 27: 43–60.

De Sutter, Gert, and Lefer, Marie-Aude. "On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach." *Perspectives* 2020, no 28(1): 1–23.

Egbert, Jesse, and Larsson, Tove, and Biber, Douglas. *Doing linguistics with a corpus: Methodological considerations for the everyday user*. [Cambridge Elements in Corpus Linguistics]. Cambridge: Cambridge University Press, 2020.

Gorshkova, Anna. "Sopostavitel'noye issledovaniye sposobov dostizheniya ekvivalentnosti v sinkhronnom i pis'mennom perevodakh." Unpublished Phd dissertation, Moskva: Rossiyskiy universitet druzhbynarodov, 2004. [Горшкова, Анна. "Сопоставительное исследование способов достижения эквивалентности в синхронном и письменном переводах." Unpublished PhD dissertation, Москва: Российский университет дружбы народов, 2004].

Grabowski, Łukasz. "Corpora and Descriptive Translation Studies." *Ways to Translation*. Bogucki, Łukasz, and Goźdź-Roszkowski, Stanisław, and Stalmaszczyk, Piotr (Eds.). Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2015: 161–188.

Gries, Stefan. "The most underused statistical method in corpus linguistics: Multilevel (and mixed-effects) models." *Corpora* 2015, no. 10(1): 95–125.

Gries, Stefan. "On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally." *Journal of Second Language Studies*, 2018, no. 1(2): 276–308.

Gurin, Il'ya. Priyemy rechevoy kompressii prisinkhronnom perevozhde s russkogo yazyka na angliyskiy. Unpublished Phd dissertation, Moskva: Moskovskiy gosudarstvennyi universitet, 2009. [Гурин, Илья. Приемы речевой компрессии при синхронном перевожде с русского языка на английский. Unpublished PhD dissertation, Москва: Московский государственный университет, 2009].

Herrmann, Berenike, and van Dalen-Oskam, Karina, and Schöch, Christoph. "Revisiting Style, a Key Concept in Literary Studies." *Journal of Literary Theory* 2015, no. 9(1): 25–52.

Holmes, James. "The Name and Nature of Translation Studies." *The Translation Studies Reader*. Venuti, Lawrence (Ed.). London: Routledge, 1988/2000: 172–185.

Ilukhin, Vladimir. *Strategii v sinkhronnom perevode*. Unpublished PhD dissertation, Moscow: 2001.

Kajzer-Wietrzny, Marta, and Grabowski, Łukasz. "Formulaicity in constrained communication: An intermodal approach." MonTI 13, 2021: 148–183.

Kalina, Sylvia. *Strategische Prozesse beim Dolmetschen: theoretische Grundlagen, empirische Fallstudien, didaktische Konsequenzen*. Tübingen: Günter Narr, 1998.

Kisiel, Anna, and Koseska-Toszewa, Violetta, and Kotsyba, Natalia, and Satoła--Staśkowiak, Joanna, and Sosnowski, Wojciech. "Polish-Bulgarian-Russian Parallel Corpus". CLARIN-PL digital repository, 2016 <http://hdl.handle.net/11321/308>.

Kluyeva, Natalia, and Bojar, Ondřej. "UMC 0.1: Czech-Russian-English Multilingual Corpus". *Proceedings of International Conference Corpus Linguistics*, October 2008: 188–195 <https://ufal.mff.cuni.cz/umc/cer/download.php?f=umc-0.1-paper-2008.pdf>.

Kopotev, Mikhail, and Mustajoki Arto, and Bonch-Osmolovskaya Anastasia (2021) "Corpora in Text-Based Russian Studies". *The Palgrave Handbook of Digital Russia Studies*. Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev (Eds). Palgrave Macmillan, Cham: 299–317 <https://doi.org/10.1007/978–3-030428556_17>.

Kotze, Haidee. "Converging what and how to find out why: An outlook on empirical translation studies." *New Empirical Perspectives on Translation and Interpreting*. Vandevoorde, Lore, and Daems, Joke, and Defrancq, Bart (Eds.). Abingdon: Routledge, 2020: 333–370.

Krasnopeeva, Ekaterina. *Lexical features of Russian translated discourse. A corpus-based comparative study of contemporary narrative prose*. Unpublished PhD dissertation: Chelyabinsk State University, 2015.

Kruger, Haidee, and Van Rooy, Bertus "Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English." *English World-Wide* 2016, 37(1): 26–57.

Kunilovskaya, Maria, and Morgoun, Natalia, and Pariy, Alexey. "Learner vs. professional translations into Russian: Lexical profiles". *Translation, and Interpreting*, 2018, no. 10(1): 33–52 <https://trans-int.org/index.php/transint/article/view/585/304>.

Kunilovskaya, Maria, and Lapshinova-Koltunski, Ekaterina. "Translationese Features as Indicators of Quality in English-Russian Human Translation." *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology* (HiT-IT 2019). Temnikova, Irina, and Orasan, Constantin, and Corpas Pastor, Gloria, and Mitkov, Ruslan (Eds.). Varna, Bulgaria, September 5–6 2019: 47–56 <https://www.aclweb.org/anthology/W19–8706.pdf>.

Kunilovskaya, Maria, and Corpas Pastor, Gloria. "Translationese and Register Variation in English-To-Russian Professional Translation." *New Perspectives on Corpus Translation Studies. New Frontiers in Translation Studies* Wang, Vincent, and Lim, Lily, and Li, Defeng (Eds.). Springer, Singapore, 2021 <https://doi.org/10.1007/978-981-16-4918-9_6>.

Kutuzov, Andrey, and Kunilovskaya Maria. "Russian Learner Translator Corpus." *Text, Speech and Dialogue*. TSD 2014. Lecture Notes in Computer Science. Sojka Petr, and Horák, Aleš, and Kopeček, Ivan, and Pala, Karel (Eds.), vol. 8655. Springer: Cham, 2014 <https://doi.org/10.1007/978-3-319-10816-2_39>.

Kutuzov, Andrey, and Kuzmenko, Elizaveta. "Two centuries in two thousand words. Neural embedding models in detecting diachronic lexical changes." *Quantitative Approaches to the Russian Language*. Kopotev, Mikhail, and Lyashevskaya, Olga, and Mustajoki, Alto (Eds.). London: Routledge, 2017.

Lanstyak, Istvan, and Heltai, Pal. "Universals in Language Contact and Translation." *Across Languages and Cultures* 2012, no. 13:1: 99–121.

Larsson, Tove, and Egbert, Jesse, and Biber, Douglas. "On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective". *Corpora* 2022, no. 17(1), forthcoming.

Laviosa, Sara. *Corpus-based translation studies: theory, findings, applications*. Amsterdam: Rodopi, 2002.

Lederer, Marianne. *La traduction simultanée. Expérience et thé*orie. Paris: Lettres modernes Minard, 1981.

Lefer, Marie-Aude. "Parallel Corpora." *A Practical Handbook of Corpus Linguistics*. Paquot, Magali, and Gries, Stefan (Eds.). Cham: Springer Nature Switzerland AG, 2020: 257–282.

Lefer, Marie-Aude. "Breaking new ground in contrastive and translation studies: Learner translation corpora to the fore". Plenary lecture delivered at the UCCTS 2021 conference in Bertinoro, Italy. 9–11 September 2021.

Łaziński, Marek, and Kuratczyk, Magdalena, and Orekhov, Boris, and Słobodjan, Elena. "The Polish-Russian Parallel Corpus and Its Application in the Linguistic Analysis." *Prace Filologiczne*, 2012, vol. LXIII: 209–218.

Łaziński, Marek, and Kuratczyk, Magdalena. "Korpus Polsko-Rosyjski Uniwersytetu Warszawskiego." *Polskojęzyczne korpusy równoległe*. Gruszczyńska, Ewa, and Leńko-Szymańska, Agnieszka (Eds), Warszawa: Wydawnictwo Uniwersytetu Warszawskiego, 2016: 83–96.

Łaziński, Marek, and Kuratczyk, Magdalena. "Pol'sko-russkiy parallel'nyy korpus Varshavskogo universiteta." *Język i Metoda* 2016, no. 3: 85–93 [Łaziński, Marek, and Kuratczyk, Magdalena. "Польско-русский параллельный корпус Варшавского университета." *Język i Metoda*, 2016, no 3: 85–93.]

Malmkjær, Kirsten. "What Happened to God and the Angels: An Exercise in Translational Stylistics". *Target*, 2003, no. 15(1): 37–58.

Massey, Gary, and Ehrensberger-Dow, Maureen. "Machine learning: Implications for translator education." *Lebende Sprachen* 62(2), 2017: 300–312.

Mikhailov, Mikhail, Tommola, Hannu, and Isolahti, Nina. "Spoken to Spoken vs. Spoken to Written: Corpus Approach to Exploring Interpreting and Subtitling." *Polibits*, 2010, no. 41: 5–9.

Mikhailov, Mikhail, and Cooper, Robert. *Corpus Linguistics for Translation and Contrastive Studies*. London: Routledge, 2016.

Mikhailov, Mikhail, and Santalahti, Miia, and Souma, Julia (Eds). *PEST. A parallel electronic corpus of state treaties*. Amsterdam: John Benjamins, 2019.

Neubert, Albrecht, and Shreve, Gregory. *Translation as Text*. Kent-London: The Kent State University Press, 1992.

NKRYA. Natsional'nyy korpus russkogoyazyka 2003–2005. Moskva: Indrik, 2005. <https://ruscorpora.ru/> [НКРЯ. Национальный корпус русского языка 2003–2005. Москва: Indrik, 2005 <https://ruscorpora.ru/>].

Olohan, Meave. *Introducing Corpora in Translation Studies*. Routledge: London, 2004.

Orekhov, Boris "Parallel'nyy korpus perevodov «Slova o polku Igoreve»: itogi i perspektivy." *Natsional'nyy korpus russkogo yazyka: 2006–2008. Novyye rezul'taty i perspektivy*. Plungyan, Vladimir (Ed.). Sankt-Peterburg: Nestor-Istoriya, 2009: 462–473 [Орехов, Борис "Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы." *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы.*

Плунгян, Владимир (Ed.). Санкт-Петербург: Нестор-История, 2009: 462–473].

Paquot, Magali, and Gries, Stefan (Eds). *A Practical Handbook of Corpus Linguistics*. Cham: Springer Nature Switzerland AG, 2020.

Pęzik, Piotr, and Ogrodniczuk, Maciej, and Przepiórkowski, Adam. *Parallel and spoken corpora in an open repository of Polish language resources. Human Language Technologies as a Challenge for Computer Science and Linguistics*. LTC Poznań 2011 <http://pelcra.pl/new/polrus>.

Piao, Zhehao, and Li, Qinghua, and Wang, Lixia. "Jiyu han-e pingxing yuliaoku de "jiu guo" dieyinci eyi guilü yanjiu." *Zhongguo eyu jiaoxue*, 2014, no 3: 46–51.

Polskoy, Sergey, and Rjéoutski, Vladislav. *The Corpus of Russian Translations*, 2021 <https://krp.dhi-moskau.org/ru/translations>.

Rybicki, Jan. "Stylometric translator attribution: do translators leave lexical traces?" *The Translator and the Computer*. Piotrowski, T., and Grabowski, Ł. (Eds). Wrocław: Wydawnictwo Wyższej Szkoły Filologicznej, 2013: 193–204.

Saldanha, Gabriela. "Translator style. Methodological considerations." *The Translator*, 2011, no. 17(1): 25–50.

Setton, Robin, and Adelina Hild. "Editors' critical foreword." *Inference and Anticipation in Simultaneous Interpreting*. Chernov, Ghelly (Ed.). Amsterdam: Benjamins, 2004: IX–XXII.

Shiryaev, Anatolij. *Sinkhronnyj Perevod*. Moscow: Voenizdat, 1979.

Shlesinger, Miriam. "Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies." *Meta* 1998, no. 43(4): 486–493.

Sitchinava, Dmitry. "Parallel Corpora within the Russian National Corpus." *Prace Filologiczne* 2012, vol. LXIII: 271–278 <https://courses.helsinki.fi/sites/default/files/course-material/4443508/CEEOL–4.pdf>.

Sönning, Lukas, and Werner, Valentin. "The replication crisis, scientific revolutions, and linguistics." *Linguistics*, 2019, no, 59 (5): 1179–1206.

Stauder, Andy, and Ustaszewski, Michal. "TransBank: A Meta-Corpus for Translation Research." *Translation in the Digital Age. Translation 4.0*. Sinner, Carsten, and Paasch-Kaiser, Christine, and Härtel, Johannes (Eds.), Newcastle upon Tyne: Cambridge Scholars Publishing, 2020: 69–80.

Szabo, Martina, and Vincze, Veronika, and Nagy, Istvan. "HunOr: A Hungarian–Russian Parallel Corpus." *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul: European Language Resources Association (ELRA), 2012: 2453–2458. <http://www.lrec-conf.org/proceedings/lrec2012/pdf/262_Paper.pdf>.

Tiedemann, Joerg. "OPUS - Parallel Corpora for Everyone." *Baltic Journal of Modern Computing* 2016, no. 4 (2). Special Issue: *Proceedings of the 19th Annual Conference of the European Association of Machine Translation (EAMT)*.

Tymoczko, Maria. "Computerized Corpora and the Future of Translation Studies." *Translators' Journal*, 1998, no. 43(4): 652–660.

Ustaszewski, Michal., and Stauder, Andy. "Corpora in Translation Studies: Why Do We Need Them and What Can Current Resources (Not) Offer to Us?." *Translation in the Digital Age. Translation 4.0*. Sinner, Carsten, and Paasch-Kaiser, Christine, and Härtel, Johannes (Eds). Newcastle upon Tyne: Cambridge Scholars Publishing, 2020: 56–68.

Ustaszewski, Michal. "Towards a machine learning approach to the analysis of indirect translation." *Translation Studies*, ahead of print, 2021: 1–19. <https://doi.org/10.1080/14781700.2021.1894226>.

Vinay, Jean-Paul, and Darbelnet, Jean. *Comparative Stylistics of French and English: A methodology for translation* (translated and edited by Juan Sager, and Marie-Josée Hamel). Amsterdam: John Benjamins, 1958/1995.

von Waldenfels, Ruprecht. "Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment." *Beiträge der Europäischen Slavistischen Linguistik: POLYSLAV 9*. Brehmer, Bernhard, and Zdanova, Vladislava, adn Zimny, Rafal (Eds.). München: Verlag Otto Sagner, 2006: 123–138.

Winter, Bodo. *Statistics for Linguists: An Introduction Using R*. London: Routledge, 2019.

Zanettin, Federico. *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing, 2012.

Ziemski, Michał, and Junczys-Dowmunt, Marcin, and Pouliquen, Bruno. "The United Nations Parallel Corpus v1.0." *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC'16)*. Slovenia: Portorož, 2016: 3530–3534.

Zimnyaya Irina, and Chernov Geliy. "Veroyatnostnoye prognozirovanye v protsesse sinkhronnogo perevoda." *Predvaritel'nyye materialy eksperimental'nykh issledovaniy po psikholingvistike*. Leont'yev, A.A., and Zhinkin, N.I., and Shakhnarovich, A.M. (Eds.). Moskva: Iya-ANSSSR, 1973: 110–116 [Зимняя Ирина, and Чернов Гелий. "Вероятностное прогнозирование в процессе синхронного перевода." Леонтьев, А.А., and Жинкин, Н.И., and Шахнарович, А.М. (Eds.). *Предварительные материалы экспериментальных исследований по психолингвистике*. Москва: Ия-АН СССР, 1973: 110–116].