Blanka FERKLOVÁ ⓘ
Palacký University in Olomouc
blanka.ferklova@upol.cz

NEW TECHNOLOGIES AND AI AS TOOLS FOR UNDERSTANDING THE KOREAS

ABSTRACT   Currently, new information about various tools using weak AI, their benefits and risks is constantly emerging. Many of the tools that can be used by Koreanists when working with texts have these technologies already firmly embedded in them. The goal of this article is to test the capabilities of several machine translation APIs, supplemented by translation using GenAI tool ChatGPT 3.5. The test was carried out mainly on a sample of South Korean and North Korean newspaper texts to identify the main problem in machine translation of Korean texts. Furthermore, the article focuses on other textual tools available to Korean scholars, such as dictionaries, corpora, and other resources provided by the National Institute of Korean language and Pusan National University tools such as the speller checker, pronunciation speller etc. and describes its potential and reliability.

Keywords: machine translation, weak AI, ChatGPT, Korean to English translation, South Korean standard language *p'yojunŏ*, North Korean standard language *munhwaŏ*, Digital Humanities, National Institute of Korean language, romanization of Korean

INTRODUCTION[1]

Nowadays, we see almost daily news about new applications, digital tools and, in particular, tools using weak artificial intelligence (AI)[2], which has been significantly publicized by the launch of OpenAi ChatGPT's large language models in November 2022. Enthusiastic voices looking for new applications and solutions are regularly replaced by skeptical and cautionary voices that point out the shortcomings and risks of technological tools using GenAI in content creation, such as using content without consent, deepfakes, a lack of understanding of the real world leading to production of inaccurate or unreliable text etc.[3] Equally, the possibilities of the beneficial use of these technologies and their incorporation into contemporary academic and working life are still being explored.

Various governmental and private organizations or professional and interest associations and individuals are trying to follow the development of technologies, experimenting with them and bringing proposals for regulations and rules or laws and, on the other hand, valuable insights into how the various tools can be used effectively. What they do agree on is that we cannot close our eyes to these developments, as these technologies and AI have already become a fixed part of many of the everyday applications we use on a daily basis, without us often realizing it.

This topic is extremely broad, and the purpose of this article is not to deal in detail with the ethical or philosophical issues of technology in our lives, nor to try to accurately capture the current state of development, which is constantly changing. This article aims to briefly introduce the most well-known technological tools, many of which also make use of weak AI, in the context of the work of a contemporary Korean scholar, Korean language learner, or expert on Korea who browses or produces a range of Korean-language texts on a daily basis, and to give a brief overview of the main benefits, but also to highlight the risks and shortcomings of these tools. I will focus on two areas in particular. First, I will analyze some of the tools used in machine translation of Korean texts, and second, I will introduce other more auxiliary tools such as converters, spell-check applications, and modern online dictionaries. In examining the machine translation of journalistic and (popular) scientific texts from Korean, the analysis will also focus on the issues of the dual Korean language standard, i.e. the South Korean standard[4]

---

[1] The transliteration in this text follows the McCune–Reischauer (MR) system. Korean names appear in their traditional order, with the family name preceding the given name.

[2] Weak artificial intelligence (weak AI) is a form designed to operate within a specific, limited domain or task. All the tools referred to as AI tools in this paper are also of this type and as Searl mentions and as we witness, despite its name, weak AI is „challenging, interesting, and difficult enough" (J.R. Searle, "Minds, Brains and Programs," *Behavioral and Brain Sciences*, vol. 3, no. 3 (1980), p. 456).

[3] Miao Fengchun, W. Holmes, "Guidance for Generative AI in Education and Research," *UNESCO*, p. 14, at https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research, 5 XII 2024.

[4] In exceptional cases, the spelling of Korean names adopted by the Korean authors themselves has been used.

pchojjuno[5] and *munhwao*[6], and use examples of translations of texts in the South Korean and North Korean standards respectively to highlight selected features as reflected or omitted by certain machine translators.

This brief analysis will conclude with an assessment of the reliability of these tools and whether they can replace the work of human translators, language teachers, or be a safe support for researchers who do not know Korean but who want to work with original Korean texts and sources. To some extent, such researchers are also students of Korean language or Korean studies who are preparing materials for their written assignments, essays, and theses, as well as their teachers who are trying to guide them well through the many pitfalls they encounter.

KOREA AND AI

The Republic of Korea is considered a technology powerhouse, contributing significantly to the rapid development of both technological devices and their 'hearts and minds' – programs based on the recently rapidly developing and highly publicized technologies based on the use of so-called weak AI.[7] It is a broad term in itself, encompassing a wide range of disciplines and applications.

In the Republic of Korea, the Ministry of Science and ICT, Artificial Intelligence Policy Division, is addressing this issue at the national policy level in its National Strategy for Artificial Intelligence *Toward AI World, beyond IT*, released at the end of 2019.[8] This strategy describes AI in the context of the so-called 4th industrial revolution, a term that has been known for many years from the concept of 'smart things', from smart phones to smart cities. The 4th industrial revolution is presented in the strategy as a process in which *human intellectual labor is replaced by AI*,[9] based also on Big Data and Hyper-Connectivity, which will eventually lead to industrial transformation, social progress, and change of life. Similarly, Minister Choi states, *Artificial intelligence, with its cognitive, learning, and reasoning abilities, is expected to assist or replace human intellectual functions in the future*.[10] Whether 'assist and replace' is the case in the field of the Korean language and its translation, this paper attempts to present and intends to be a contemporary probe into the possibilities and limits that AI has significantly

---

[5]   표준어, lit. Standart language (of the Republic of Korea).

[6]   문화어, lit. Cultural language (of the Democratic Peoples Republic of Korea).

[7]   The most discussed of which is currently the so-called Gen AI, Generative Artiftial Intelligence, that automatically generates content in response to prompts written in conversation interfaces. Its best known current representative is the aforementioned ChatGpt, based on a large language model (Miao Fengchun, W. Holmes, "Guidance...," p. 8).

[8]   That is, almost 2 years before the first launch of the revised Chat GPT from Open.ai in November 2022.

[9]   "National Strategy for Artificial Intelligence," *Ministry of Science and ICT*, p. 6, at https://www.msit.go.kr/bbs/view.do?sCode=eng&nttSeqNo=9&bbsSeqNo=46&mId=10&mPid=9, 5 XII 2024.

[10]   Ibid.

influenced. In this context, it is also worth mentioning North Korea's current interest in AI and its applications as described by Lim.[11]

In addition to the national level, the activity of large South Korean companies[12] is also evident in international cooperation with global industry leaders. For example, Meta's owner Mark Zuckerberg's recent visit to LG and Samsung in February 2024[13] or reports of a development of an internal app similar to ChatGPT at Samsung in collaboration with Naver.[14] This decision is said to be prompted by unfortunate incidents where Samsung engineers inadvertently leaked confidential company information using ChatGPT, which again points to some of the risks of using these technologies. As a prevention Samsung has imposed a ban on the use of external generative AI models by its employees and may replace it with its own secure application.

However, the use of machine artificial intelligence is not new in recent years. Many of the tools that use weak AI that we will mention here, whether Korean or of other origin, are older and are often based on many years or decades of development:

MACHINE TRANSLATION OF KOREAN TEXT

The technologies used in Korean language translation have been the focus of a number of studies that have taken a holistic view, as well as specifically addressing individual issues in machine translation of Korean,[15] analysis of translator errors,[16] specifically the use of artificial intelligence in translation,[17] etc. These works build on decades of

---

[11] Lim Tai Wei, "North Korea's Artificial Intelligence (A.I.) Program," *North Korean Review*, vol. 15, no. 2 (2019), pp. 97-103.

[12] J. Jung, "Top 6 Gen AI Projects from South Korean Tech Giants," *KoreaTech Today*, 31 X 2023, at https://www.koreatechtoday.com/a-glimpse-into-2023-top-5-gen-ai-projects-from-south-korean-tech-giants/, 5 XII 2024.

[13] Park Minseo, "Zuckerberg's South Korea Visit: Collaborating on AI and XR Technologies with LG and Samsung," *KoreaTech Today*, 28 II 2024, at https://www.koreatechtoday.com/zuckerbergs-south-korea-visit-collaborating-on-ai-and-xr-technologies-with-lg-and-samsung/, 5 XII 2024.

[14] Kyung Mi, "Samsung Develops In-House AI Platform for Employees, Avoiding Leaks and Enhancing Data Privacy," *KoreaTech Today*, 16 V 2023, at https://www.koreatechtoday.com/samsung-develops-in-house-ai-platform-for-employees-avoiding-leaks-and-enhancing-data-privacy/, 5 XII 2024.

[15] J. Lewis, "Globalization, Culture and Technology: The Case of Korean Machine Translation," *Hitotsubashi Journal of Social Studies*, vol. 35, no. 1 (2003); Lim Soon Jeung, Mi-sun Han, "Kugŭl Pŏnyŏgŭi Saryeyŏn-gu: Pŏnyŏkp'umjilgwa Pŏnyŏktanwiwaŭi Kwan'gyerŭl Chungshimŭ-ro" (A Case Study on "Google Translate: The Correlation between Translation Quality and Units of Translation in Machine Translation), *Beonyeokagyeongu*, vol. 15, no. 1 (2014); Park Junsuk, "A Comparative Study of AI translation – Google Translate and Papago," *Studies in Linguistics*, vol. 65 (2022).

[16] Park Junsuk, "Analysis of the Types of Translation Errors by the Google Translator," *The Jungang Journal of English Language and Literature*, vol. 59, no. 4 (2017); Lee Jeong-Hwa, Cha Kyung-Whan, "An Analysis of Korean-English Translation Errors in Google Translate," *The Journal of Linguistic Science*, no. 89 (2019).

[17] Noh-shin Lee, Sinjae Lee, Jaeyeong Lee, Juhui Lee, "Tongbeonyeogui Miraejipyeong: Ingongjineunggwa Sotonghyeong Yunghaptongbeonyeong Yeongu" (Future Horizons of Translation & Interpretation: Artificial Intelligence and Interactive-Converged Translation & Interpretation),

research and improvement in machine translation, which was stimulated in the late 1940s by the initiatives of Warren Weaver, and has been followed up by Bar-Hiller's research since the 1950s. In his 1953 study Bar-Hiller pointed out general problems of machine translation, especially idioms and syntactic categories and the generalization of language description.[18] By the 1960s a machine translation company SYSTRAN was founded, and is still active today.[19]

With the development of machine translation, especially rapid since the 1990s, the specific issues in translation have become increasingly known and described; both for different languages, based on descriptions of their specificities in general, and for translation between two certain languages, based on their mutual differences. In a similar way, for example, the translatologist Anthony Pym presents machine translation between English and Catalan[20] and the specific problems posed by the different grammatical categories.[21] In this paper, it will be assessed how the issue was manifested in the analyzed translations, i.e. what accuracy translators achieve and whether they can possibly replace a human translator.[22]

In the following overview, we present the main problems that regularly appear or have been manifested in the translation of Korean text into English in this probe. It is a combination of both universal and typically Korean elements, and we can assume that the problems in translating from Korean are roughly as follows:[23]

Idioms and phrases are mentioned as early as the 1950s by Bar-Hiller.[24] Phrasemes may or may not occur in isolation, in complete form, their meaning and boundaries within a given text are usually not marked graphically. In the texts analyzed in this research, phrasemes occurred to a lesser extent, only in newspaper texts, which, in contrast to scientific texts, use more phrases and other means to diversify and update the content and make reading more enjoyable.

Beonyeokagyeongu, vol. 17, no. 2 (2016), pp. 65-89; Ahn Miyoung, "MT Problems and Its MTPE Suggestion with Regard to Structural Differences between English and Korean," *The Journal of Mirae English Language and Literature*, vol. 25, no. 1 (2020).

[18] Y. Bar-Hillel, "Some Linguistic Problems Connected with Machine Translation," *Philosophy of Science*, vol. 20, no. 3 (1953), pp. 217-225.

[19] Systran provides a web API that allows free use. The list of languages that can be translated via English includes Korean and more than 40 other languages, so it was also included in the testing (https://www.systransoft.com/lp/korean-translation/), 5 XII 2024.

[20] A. Pym, "AI and the Future of Translation: In Search of Evidence," *Youtube*, 18 VII 2023, at https://youtu.be/46rUOOmNlkA?si=s8uGcDkIneM5fBtd. In addition to studies, Pym has recently published a number of videos in which he addresses the issues of machine translation and AI.

[21] On problems of gramatical cathegorie in the perspective of a machine translation see D. Zeman, *The World of Tokens, Tags and Trees*, Prague 2018, Studies in Computational and Theoretical Linguistics.

[22] Anthony Pym uses the term "singularity" for this phenomenon, where the quality of machine translation with AI is equal to that of a human translator, while of course machine translation is incomparably faster. The term refers to Ray Kurtzweil's book *The Singularity is Near* (2005) and refers to AI in general (A. Pym, "GPT in the Training of Translators," *Youtube*, 27 IV 2023, at https://www.youtube.com/watch?v=b9U_FUaneso, 0:20, 5 XII 2024.

[23] This list makes no claim to completeness and presents only roughly the basic issues.

[24] Y. Bar-Hillel, "Some Linguistic Problems…," p. 217.

Specific vocabulary such as terms[25], neologisms and words written in Chinese characters. Special terminology covers a very large part of the vocabulary in most languages, and in Korean there is also considerable homonymy of terms due to their frequent Sino-Korean origin, where the correct meaning of a word is often only apparent from the Chinese characters used to clarify the meaning. In texts analyzed in this paper, several linguistic terms appeared and their translation was analyzed. Neologisms, due to their innovative nature, are recorded in dictionaries with a certain delay[26], and therefore their identification and successful translation depends on the degree of their integration in the dictionary or other source used by a given translator.

Names of persons, organizations, and local names etc. It is important to identify them in the text, i.e. as a first step, to recognize that they are such names. Korean does not distinguish between upper and lower case, so names are not marked in any way. The exception is North Korean texts, where the names of the leaders of the country are usually given in larger or otherwise emphasized letters. Names of people or families, organizations and local or geographical names, are not marked, and therefore their identification is a matter of correctly identifying the complex context of the text.

Polysemous words and homonyms – translation of polysemous words and homonyms was also an expected problem, as it may be difficult even for human translators and is very context-sensitive due to its ambiguity.

From a more grammatical perspective on translation, the expected problem is the grammatical **gender** and grammatical **number**, especially in sentences in which the subject is not expressed or is not gender-specific (in the case of persons). On the other hand, Korean's complicated politeness system causes no difficulties in translating from Korean into English, unlike when translated in reverse, i.e. from English into Korean.[27]

Syntactic structures present a large number of problems, e.g., in the case of the use of the Korean zero case marker[28], the correct recognition of the actants in a sentence. Machine Translators also consider different lengths of text, which can cause difficulties for very long sentences or complex sentences. A classical problem are coordination phrases, where it is not always clear what the coordination phrase is related to, whether just a particular word or the whole phrase.[29]

---

[25] In this category we can also include culturally specific words that are difficult to translate because they have no exact equivalent in the target language, e.g. *kimchi*, *hanok*, *hanbok*, etc., which, however, occurred minimally in the analysed texts.

[26] Nam Kilim, Lee Soojin, Jung Hae-Yun, "The Korean Neologism Investigation Project: Current Status and Key Issues," *Dictionaries: Journal of the Dictionary Society of North America*, vol. 41, no. 1 (2020), pp. 105–129.

[27] The possibilities of translating politeness and mutual equivalence of polite/impolite means, specifically between Czech and Korean, have been shown in the work of Kwak Baštanová (Y. Kwak Baštanová, "Grammatical Expressions of Politeness in Czech and Korean: Tykání and Vykání in Korean," PhD diss., Palacký University Olomouc, 2021).

[28] When a given word in a sentence is not marked by a grammatical device – an ending or particle, and its function therefore results from the context (V. Pucek, *Gramatika korejského jazyka*, Praha 2005, p. 215).

[29] O. Bojar, *Čeština a strojový překlad*, Prague 2017, p. 10, Studies in Computational and Theoretical Linguistics, 11.

MACHINE TRANSLATORS OF TODAY

There are dozens of language-specific translators and lists of machine translators to be found on the internet today.[30] Of course, the highest number of translators is listed for English (based on our source, there are 58 translators), but Korean also reaches a high number of 42 machine translation APIs. This number and coverage of languages is likely to change in the future, whether new applications are developed, existing ones increase the range of languages, or developments are different. However, many of these translators are commercially based and are not freely available or require registration. For this article, therefore, only a few of the most well-known, freely available translators were used for the analysis. These are Korea's two largest translators, Naver Papago (NP)[31] and Daum Translate (DaT)[32], popular Google Translate (GT)[33], DeepL Translate (DeT)[34] and Microsoft Bink (MB)[35], and of the lesser-known but readily available ones, the oldest, Systran (SYS)[36], and finally, Promt (PRO)[37]. The last tool that was included in the analysis is the aforementioned ChatGPT 3.5 (CG)[38], which, although not originally intended as a translation tool, its multilingualism allows for this function and use. As a language model, ChatGPT stands out in particular for producing convincingly stylized grammatically correct or plausible sounding sentences, suggesting promising potential for use as a translation aid. All listed machine translators were used in the unpaid, public domain version. While a number of translators allow for paid member use, which brings many advantages in individualizing the translator to the user's needs, this would bias the results of this analysis.

ChatGPT creates linguistic content from the sources it obtains in many languages, especially on the Internet, and other translators also make use of the vast amount of online text. In terms of the representation of languages on the Internet, research shows that English language websites take up the largest share (accounting for about half of the texts), followed by languages such as Chinese, Spanish, Russian and also

---

[30] Available at eg. https://machinetranslate.org/, 5 XII 2024.

[31] https://papago.naver.com/?sk=ko&tk=en, neural machine translation, 5 XII 2024.

[32] https://translate.kakao.com/, neural machine translation, 5 XII 2024.

[33] https://translate.google.com/?sl=ko&tl=en&op=translate, statistical and neural machine translation, 5 XII 2024.

[34] https://www.deepl.com/en/translator, neural machine translation, 5 XII 2024.

[35] https://www.bing.com/TRANSLATOR?toWww=1&redig=38C889F5CA1D4E959DD147B3B-4C885DD, statistical and neural machine translation, 5 XII 2024.

[36] https://www.systransoft.com/lp/korean-translation/, hybrid, rule-based, statistical and neural machine translator, 5 XII 2024.

[37] https://www.online-translator.com/translation/korean-english, 5 XII 2024. PROMT is hybrid, rule-based, statistical and neural machine translator.

[38] https://chat.openai.com/, 5 XII 2024, newly appeared also Translate GPT application (https://chatgpt.com/g/g-5bNPpaVZytranslate-gpt, 5 XII 2024). It does not rely primarily on dictionaries, but on large amounts of text data.

Japanese.[39] In the case of Korean, however, no studies have yet taken into account the fact that websites are overwhelmingly made in the Republic of Korea and thus texts with spelling and vocabulary specific for the Democratic People's Republic of Korea will be significantly fewer. Hence, the hypothesis is that web translators are likely to be lacking it. Therefore, this analysis includes the translation of North Korean newspaper texts with certain specific phenomena to test this hypothesis.

The results of the translations were judged according to the criteria of the clarity of the output and its accuracy in relation to the original meaning; in the case of the popular scientific texts the official translation in English was also available. Anthony Pym[40] introduces automated metrics for evaluation of machine translation such as the BLEU. system for measuring translation accuracy, but also favors a reader-based approach, i.e., a subjectivized reader evaluation of translations, which in its simplified form is adopted in evaluating our experiments. Thus, we use the evaluation of translation accuracy within the context of the sentence and text, and in particular focus on the assessment of errors and their types.

For machine translation, contemporary news texts with diverse topics were selected from South Korean online newspapers, namely Hankyore, Naver News, and North Korean Rodong Sinmun as well, with diverse topics. In addition, the analysis covers also a translation of a chapter from the National Institute of Korean Language's popular educational publication *Urimal Imojŏmo* (*Everything You Want to Know about Korean Language*), Chapter 5 of *Nanmarŭi se jongnyu* (*Three Types of Words*) was chosen. Unlike newspaper texts, there are scientific terms, especially of Sino-Korean origin, and also words written in Chinese characters.

A PROBE INTO MACHINE TRANSLATION FROM KOREAN INTO ENGLISH WITH SEVEN MACHINE TRANSLATORS AND CHAT GPT

For this analysis, which intends to test the translation abilities of a few selected machine translators, samples of newspaper and popular scientific texts were chosen. Previous studies dealing with machine translation of Korean have tended to focus on specific aspects of the translation of a particular grammatical phenomenon, often polysemous, as well as analyses of the solutions provided by the translators,[41] or comparisons between

---

[39]   D. Pimienta, D. Prado, A. Bianco, „Twelve Years of Measuring Linguistic Diversity in the Internet: Balance and Perspectives," *UNESCO*, at https://unesdoc.unesco.org/ark:/48223/pf0000187016, 5 XII 2024. Twelve years of measuring linguistic diversity in the Internet: balance and perspectives. W3Techs estimations are as follows English: 50,8%, Spanish (2nd) 5,7%, German 5,1%, Japanese 4,6%, French + Russian 4,3%, Polish 1,7%, Chinese 1,3%, Korean and Czech less than 1% (Czech 0,9%, Korean 0,8%). https://w3techs.com/technologies/overview/content_language, 5 XII 2024.

[40]   A. Pym, Exploring Translation Theories, London–New York 2023.

[41]   Lee Jeong-Hwa, Cha Kyung-Whan, "An Analysis…"; Park Junsuk, "A Comparative Study of AI translation – Google Translate and Papago," *Studies in Linguistics*, vol. 65 (2022); Lim Soon Jeung, "Kugŭl shin'gyŏngmang pŏnyŏg ŏlmana chinhwahaenna? Ch'och'anggiwa hyŏnjaeŭi pŏnyŏg p'umjirŭl

the translators and human translators[42] or the translation of very specific features such as irony.[43] This analysis will go a little wider in the number of texts and translators, which will be less in-depth, but will show and compare problematic areas.

As for the translation of newspaper texts, most of the largest South Korean newspapers[44] provide an English version, which is usually easily available online. However, not all news is translated, as one of Korea's largest newspapers, the *Chosun Ilbo*[45] states on its website. There are therefore a large number of journalistic texts that are not accessible without knowledge of Korean, but which contain potentially important information for understanding Korean society and the current political situation or other issues.

For this analysis, news from the Hankyoreh newspaper and Naver News were selected as well as popular scientific texts from the publication on Korean language *Urimal Imojŏmo* (*Everything You Want to Know About the Korean Language*). Other texts were selected from the North Korean Rodong Sinmun. As mentioned above, the different orthographic norms of the DPRK, to some extent also vocabulary and idioms, as well as morphological differences, may pose a problem for the translation of these texts. It is in a sense a specific variety of Korean, similar to some extent to dialects, but far superior to them in meaning and number of speakers and texts.

The tested texts were selected so that they contain idioms, homonyms, personal, local and organizational names, as well as neologisms. A number of these phenomena can be illustrated by the headline of an article from the Hankyoreh,[46] criticizing the organization of the global scouting camp Jamboree that took place in the summer of 2023 in Korea, which reads very loosely translated as 'The problem are incompetent civil servants – why the Jamboree went wrong'.

Here we find the neologism *ŏgong*, which is enclosed in quotation marks to indicate expressiveness or unusualness in the original Korean text. It is a blend of the words *ŏtchŏda kongmuwŏn* (who happened to be a civil servant)[47] to refer to professionals who are civil

---

chungshimŭro" (What Progress Has Been Made in Neural Machine Translation: The Case of Google Translate), *P'ŭrangsŭŏgwŏn munhwayesuryŏn'gu*, vol. 83 (2023).

42   E.g., Lee Chang-Soo, "Ch'aetGPT Ch'urhyŏn Ihu Kigye Pŏnyŏkkwa In>gan Pŏn Station Kanŭi Pŏn Station Munch>e Ch>ai Pyŏnhwa Yŏn-gu" (A Follow-up Study of Stylistic Differences between Human and Machine Translation with ChatGPT Added in the Mix), *Beonyeokagyeongu*, vol. 24, no. 3 (2023), performed on newspaper texts.

43   Park Soojung, Eunsil Choi, "Ch'aetGPTŬi Airŏni Pŏnyŏk Hwaryong Kanŭngsŏng Koch'al" (A Study of Translatability of Irony in ChatGPT), *Beonyeokagyeongu*, vol. 24, no. 2 (2023).

44   Like Chosun Ilbo, Hankyoreh, Jungang Ilbo etc.

45   „…We make selections from the evening edition and carry only main stories, such as the Editorials, and top business, politics and national news." "Q/A," *Choson Ilbo*, at https://news.chosun.com/english/faq.html#:~:text=should%20be%20followed.-,7.,business%2C%20politics%20and%20national%20news, 5 XII 2024.

46   Hanyong Sŏng, "Munjenŭn Ŏgongida...chaembŏri-ga sanŭ-ro kan iyu" (The Problem Are Incompetent Civil Servants – Why Jamboree Why Did Go Wrong), *Hangyoreh*, 20 VIII 2023, at https://www.hani.co.kr/arti/politics/politics_general/1104943.html, 5 XII 2024.

47   '어쩌다 공무원' It mainly criticaly refers to professional, fixed-term civil servants who have not passed the civil service examination.

servants without having passed the civil service examination.[48] While a number of machine translators have attempted to translate the individual syllables of the term as 'fishing ball'[49] (GT), 'fish ball' (DeT) or just 'fishing' (NP, SYS, PRO), 'fisherman'[50] (MB), or in one case the translator has omitted the translation altogether (DaT) or using only the romanization 'Eo-gong' (CG).[51] Thus, none of the translators really translated the meaning of the neologism. Surprisingly, sentences where the translation of the term was omitted sounded more meaningful in the end than those where it was translated incorrectly.

The other problematic item in the headline: the name of the event *chaembŏri* (*Jamboree*) was translated almost in all cases correctly, only Chat GPT did not translate it and only substituted a romanized guess with 'Jamvery'. The second part of the headline contains part of an idiom [52], which refers to the organization of the event, with too many authorities involved in the preparation to ultimately move in the right direction. In a test of recognition of this idiom, all translators failed consistently, translating only the literal meaning of 'went (to the) mountains', with only Daum translating 'go mountains' and thus ignoring the past tense expressed in the verb form. This translation is not even remotely indicative of the true meaning of the idiom and can be judged insufficient.

The second article by Naver News,[53] which dealt with the Democratic Party's campaign promises was translated in many cases very lucidly. The worst result in terms of readability was clearly the Promt translator. All translators were able to identify the name of the political party, even though it appeared in two different forms, i.e. *Minjudang* and *Tŏburŏ Minjudang* (*Democratic Party* and *Together Democratic Party*).[54] Also, the name of the politician who speaks in the article was identified and romanized identically in almost all translations. Except for Promt, they also used the same transliteration using Revised Romanization, only Promt translated the name *O* surprisingly as *Au* and not the more common Oh.[55] Promt also translated some words as if they were names, such as *sŏmin* (*citizen*), which appeared as Seo-min, and even the grammatical construction *-kil paranda* (*to wish st.*) as the name Gil Baranda. The geographical

---

48　Referred to as a term used in the press (https://ko.dict.naver.com/#/entry/koko/2895fe87e7664aba-92c645a9cf8f1559, 5 XII 2024).

49　Ŏ being analyzed as „fish" referring to a Sino-Korean morpheme 魚, *gong* as the pure Korean word „ball".

50　工 *gong* – work, somebody who is doing particular work.

51　This translation partly indicates that it is a (personal) name also due to the use of a capital letter at the beginning of the word.

52　Translates like: Too many cooks spoil the broth. *Sagongi manŭmyeon paega sanŭro kanda* (사공이 많으면 배가 산으로 간다), lit. if there are many shipmasters, the ship will go to the mountains.

53　Jiwŏn Hŏ, "Minjudang 'Pulbŏpchuch'a Ch'ŏbŏl Nanimch'iryo Yugŭp Hyu-ga Hwaktae' tŭng Kongyak" (Democratic Party Pledges to Punish Illegal Parking and Expand Paid Leave for Infertility Treatment), *Naver News*, 4 II 2024, at https://n.news.naver.com/mnews/article/079/0003880864, 5 XII 2024.

54　https://theminjoo.kr/main/, 5 XII 2024.

55　Considering Promt is a Russian tool, we may presume that this difference may occure due to use Russian as base for translation, however according to rules of cyrillization of Korean, the letter is still transcribed as "o".

names that appeared in the article (Yeouido, Seoul) were correctly identified by all translators and translated in a completely identical way. Thus, most of the tested translators had no difficulty in converting the names, even though they are not capitalized in Korean as in Latin, and were very accurate, except for the above mentioned Promt.

Notable in this article was the way in which the issue of nominal gender, or the respective personal pronouns, was addressed. The politician who speaks in the article is identified only by the gender-neutral word *minjudangwŏn* (a member of the Democratic Party). Korean names can also be considered gender-neutral[56], and therefore, it was not possible to tell from the plain text whether the politician in question is male or female. However, the machine translators mostly opted for the additional title of 'Mr'. Only the translator Daum was able to avoid the title and used only a gender-neutral name, however, in the following text he inevitably uses the masculine pronoun 'he' and thus also defines the gender of the person as male.

As a side finding, the different ways of completing the information regarding the date emerged from this analysis. It is a peculiarity of South Korean newspaper news that when they report an event, they do not give the month it took place, only the day. It is then obvious that it is the month when the news was published. Daum was the only translator to correctly complete the date as is usual in English, although the information was not provided to the translator. However, it is questionable whether this was just a coincidental match.

We can summarize that the analysis of the machine translations of news articles has confirmed the problems of translating terms, neologism, names, idioms, as well as the specific problems of gender marking that arise when persons are named in the texts. As a side-effect, the addition of the date of the event described in the text was observed. The overall intelligibility of the text varied, but it can be said that the best results were achieved by the DeT, GT and NP translators.

For the comparison of translations of a popular scientific text, two parts from *Urimal Imojŏmo* (*Everything You Want to Know About the Korean Language*)[57] published by the National Institute of Korean language, chapter 5, *Nanmarŭi se jongnyu* (*Three Types of Words*) were selected. From it, two sections were collated, one presenting linguistic terms and the other giving examples of Sino-Korean vocabulary that were written in Chinese characters. The results of the machine translators were very different in these cases. In particular, the translation of the terms and the subsequent intelligibility of the text with these translations were monitored, as well as the way in which the Chinese characters were translated.

Both Korean translators NP and DaP did not translate the basic linguistic term *hanjaŏ* (Sino-Korean word),[58] which was the main topic of the text, and translated it incorrectly as 'chinese characters'. Naver Papago, which, unlike other translators, also

---

[56] Although in Korean there is a tendency to use certain names more for girls and others more for boys, it cannot be decided that the bearer of a certain name can be both female and male.

[57] 우리말의 이모저모, 3 장 낱말의 세 종류. Available at https://www.korean.go.kr/front/etcData/etcDataView.do?mn_id=46&etc_seq=183&pageIndex=14, 5 XII 2024.

[58] 한자어.

displays the used vocabulary in the form of dictionary entries below the translation windows, surprisingly stated the correct meaning 'Sino-Korean word' in this section, but it was not used in the translation above. For the other terms *oeraeŏ* (loanword)[59] and *koyuŏ* (Korean native word)[60] which together form the basic three groups of Korean vocabulary according to its origin, the NP translates completely incorrectly as 'foreign/native languages', which completely loses the meaning of the text. Other machine translators gave the translations of the last two terms very inconsistently, often with different translations in the text, e.g. as 'Western foreign words, western loanwords' (GT), 'hanja words, kanji' (DeT), completely wrong with 'Western outsiders, indigenous words' (PRO). Chat GPT avoided translation entirely and kept the terms in the original Korean form, not even romanized.

Chinese characters which appeared in the text were translated unsystematically by both Korean NP and DaT translators, i.e. their meaning was not always recognized. Daum translated only one and replaced the rest with a series of question marks. The Chinese characters were translated or transcribed by the GT and DeT translators quite successfully,[61] Microsoft Bing provided translations of the meaning of the listed character words as sometimes consisting of only partial and not all characters, e.g. *paek-il-chang* (essay writing competition)[62] translates only from the first two characters as *paek-il*, 'hundred days'.

The two lesser-known translators Systran and Promt had the worst results in terms of accuracy and comprehensibility in translation of Chinese characters. Some solutions could be characterized as *ridiculous mistakes from the times of the beginning of machine translations*, as mentioned by Bojar.[63] Systran converts Sino-Korean words inappropriately as Chinese characters, but the Western foreign words are quite usable. It is the only translator which correctly recognized the aforementioned *paek-il-chang* (essay writing competition) as essay writing, but didn't translate other relatively easy Sino-Korean words written in Chinese characters, such as *myŏndo* (shaving). Systran was the only translator to miss the context of the sentence that words of foreign origin came with spoken speech, while Sino-Korean vocabulary entered mostly with written texts. The expression *mallo* (in spoken language) was translated as 'with horses', which is homonymous with it, but completely wrong.

The other translator with the worst result: the Promt translator, was the only one that significantly reduced the translated text to roughly one half, and its translation did not make much sense. It converted Chinese characters in the same way as other translators, i.e. by translation or transliteration, however, surprisingly, it lists the last three characters in the Greek alphabet.

---

59   외래어.

60   고유어.

61   However, they left the last set of characters completely untranslated, which was the correct solution given the context. However, Deepl also added the translation of one character (or stated the same one twice), which, compared to the original, means an initiative expansion, which is erroneous.

62   백일장.

63   O. Bojar, *Čeština*..., p. 10.

As for the translation of texts containing terms and Chinese characters, it can be said that all translators had problems and made a lot of mistakes. The examples that were given in Chinese characters were partially translated, usually some of them incorrectly, Daum almost completely avoided the translation of the characters and replaced them only with question marks. Chat GPT also avoided the translation of Chinese characters and kept all terms and Chinese characters in their original form, not modifying them in any way. This approach is partially acceptable and applicable, for example, to Korean students who can read Hangul and have an awareness of Chinese characters, because the translation was otherwise very precise.

As a final addition to this brief comparison of the capabilities of the most used machine translators between Korean and English, we focused on newspaper texts from the North Korean *Rodong Sinmun*.[64] The aim was to verify the possibilities of translating words in North Korean spelling, which particularly concerned words with the initial 'L', which is one of the most significant spelling differences from the South Korean standard. Furthermore, to verify translation of possible different vocabulary, specific idioms and certain morphological differences that occurred in the texts were analyzed.

When comparing translations of North Korean texts, the machine translators have very different results from each other. Nevertheless, many of them were able to identify and correctly translate words written according to the North Korean spelling standard, i.e. they mention words with the initial 'L'. Also, specific terminology not used in South Korea was translated correctly by some of the translators.

The best results in the translation of North Korean terms and words in North Korean spelling went to Deepl Translate, which translated all the tested items correctly. The least successful and practically unintelligible was the translator Promt, which did not translate any of the monitored expressions correctly.

Korean translators Naver Papago and Daum translated some terms with North Korean spelling only by romanization, e.g. *raengdong* (freeze) as 'lanngdong'. GT, MB and CG correctly identified words with the North Korean spelling, but chose the translation 'refrigeration' more often, except for GT. Another example is *laeil* (tomorrow), which was only indirectly translated as 'future', or its translation was omitted.

Morphological differences from the standard South Korean were rarely found in the texts, for example in the forms of the verb *toeda* (to become).[65] None of the translators had any trouble identifying it and all were translated correctly, as this difference was probably evaluated at the level of a common typo, and thus the form was not difficult to be identified correctly.

---

[64] Two articles were analyzed: Hyosim Kang, "Widaehan Ŏmŏniŭi Sarangŭl Chŏtchulgi-ro Hayŏ Chogugŭi Mirae-ga Murŏngmurŏk Charanda" (With the Love of a Great Mother as the Source of Milk, the Future of Our Country Grows Prosperously), *Rodong Sinmun*, 4 II 2024, at http://www.rodong.rep.kp/ko/index.php?QEBAQEAx, 5 XII 2024; and "Kyŏngaehanŭn Kimjŏngŭn dongjikkesŏ Kimhwa-gun chibang gongŏp kongjangdŭrŭl hyŏnji jidohashiyŏtta" (Respected Comrade Kim Jong-un Provided On-site Guidance to Local Industrial Factories in Kimhwagun), *Rodong Sinmun*, 4 II 2024, at http://www.rodong.rep.kp/ko/index.php?QEBAQEAx, 5 XII 2024.

[65] North Korean form *toeyŏya, toeyŏtta* (must become, became) instead of South Korean standard *toeŏtta, toeŏya* (되어야/ 되여야, 되었다 /되였다).

In the texts, the designation for the North Korean Labor Party, Rodongdang, transposed by the expression *Ŏmŏnidang* (Mother's Party), or *Ŏmŏni* (mother) appeared. This is a typical manifestation of North Korean journalistic rhetoric style. None of the tested tools translated the expression figuratively, but only literally, but they differed on whether it was 'great mother' as stated by most translators, or 'great love' (Chat GPT), when the translation 'mother' was completely omitted. The last note belongs to the translator Promt, who is the only one who completely omits in the translation of a headline of the second analyzed article the word *sarang* (love), and surprisingly translates it as 'mourning', which makes the title completely lose its meaning, while the other translators, even with the above-mentioned difficulties, managed to convey the meaning quite satisfactorily.

Of these, the translation by Deepl also stands out in the sense that it translates the sentence into the past tense, even though it refers more to the present and the future. Overall, however, it can be said that apart from no translator identifying *Ŏmŏni* (mother – the Labor Party), the translators did a pretty good job of meaningful translating the rest of the text.

However, the overall evaluation of the translations of North Korean texts is not very positive. Words containing North Korean spelling or a North Korean word were often a stumbling block, surprisingly especially for the Korean translators Naver Papago and Daum, that in many cases provided the words only as their romanization, which was also not uniform. In contrast, minor morphological differences in North and South Korean were translated flawlessly. The phraseological use of the word 'mother' in the sense of Labor Party was not recognized by any of the translators.

On the sidelines of the evaluation of machine translation, one more problem should be mentioned, which is simply copying and pasting of the texts into machine translators. Newspapers and other tested texts that when transferred to translators, contained the original typesetting. It often divided sentences into more units which the translators tended to translate separately, also in some cases the gaps between words were eliminated which also made successful translation difficult or impossible. Careful editing and checking of the translated text before translation is therefore very important. We mention these simple mechanical obstacles because they caused absurd mistakes in the translation, the meaning of which is thus lost. Although the translation itself is done immediately by the machine translator, pre-editing and post-editing, i.e. additional work with the text is often necessary and requires a considerable amount of time.

DIGITAL TOOLS OF THE NATIONAL INSTITUTE OF KOREAN LANGUAGE AND PUSAN NATIONAL UNIVERSITY

The creation of various digital tools came hand in hand with the development of computerization and digitization. The path to the development of Korean digital resources, which is a prerequisite for the current dynamically developing field of Digital Humanities, from the digitization of historical materials and the continued government

support of many institutions engaged in digitization, is mentioned by the prominent researcher in the field of Korean Digital Humanities Javier Cha.[66] Its roots can already be seen in the 1990s and especially in the period after the Asian financial crisis, where released workers in particular were transferred to state projects, including the digitization of historical resources.[67] This work continues to this day and in many respects also relates to work on Korean language resources.

Nowadays, the National Institute of Korean language, as the representative body of national language policy and language research, offers a large number of digital resources and applications that can be used by public users and researchers. A significant contribution was received by the National Institute of Korean Language in 2017 in the amount of 17.5 million USD for a period of 5 years in order *to create 15.5 million bags of words representing the modern Korean language for AI-driven linguistic analysis*[68]. A major step was the digitization and online launch of the Standard Korean Language Dictionary in 2002,[69] which is adopted by major web search engines providing dictionaries such as Naver or Daum.

Very significantly, especially in the area of neologisms, this representative dictionary is complemented by the modern digital dictionary *Urimalsaem* (The Source of Our Language)[70], which among other sources captures neologisms that are inserted by the public users[71] and edited by experts. It is therefore an online, open source, and collaborative dictionary[72] which was first launched in 2016. Open source presents certain risks, e.g. errors and insertion of individual coinage, which must be edited and a mechanism created to identify and edit or remove such expressions from the dictionary.[73] We men-

---

[66] J. Cha, "Digital Korean Studies: Recent Advances and New Frontiers," *Digital Library Perspectives*, vol. 34, no. 3 (2018); J. Cha, "The Digital Turn in Korean Studies: Trends, Challenges, and Prospects," *Youtube*, 15 IX 2021, at https://youtu.be/sheRZim_wwc?si=ZafNJJVFez8TbK_N, 5 XII 2024.

[67] J. Cha, "The Digital Turn..."

[68] I Ung, "In'gongjinŭngyong Han'gugŏ malmungch'i 155 Ŏgŏjŏl kuch'uk 5 nyŏn'gan 175Ŏk Chiwŏn" (Establishment of 15.5 Billion Korean Words for Artificial Intelligence... 17.5 Billion Won in Support over 5 Years), *Yonhap News*, 9 IX 2017, at https://www.yna.co.kr/view/AKR20171008048600005, 5 XII 2024.

[69] 표준국어대사전, https://stdict.korean.go.kr/main/main.do, 5 XII 2024. The compilation of Standard Korean Language Dictionary was commenced in 1992, the dictionary's first edition was published in 1999 and on the compact disc in 2001. The online dictionary was launched in 2002 and revised in 2008 (https://www.korean.go.kr/front_eng/intro/intro_03.do, 5 XII 2024 National Institute of Korean language – History). Standard Korean Language Dictionary also contains the standard pronunciation in Korean notation and is therefore a reference when determining the correct pronunciation.

[70] 우리말샘, https://opendict.korean.go.kr/dictionary/, 5 XII 2024.

[71] Similar to how e.g. the Wikipedia, internet encyclopedia, works.

[72] Nam Kilim, Lee Soojin, Jung Hae-Yun, "The Korean Neologism..."

[73] The paper states that since 2015 the data is obtained mainly from the Web and the threshold frequency has been increased from 1 to 3 occurrences from the following to reduce errors and individual coinage. However, due to the recycling nature of a number of texts and their placement without much change in other places on the web, quite unique expressions can still enter the dictionary. The article therefore describes other ingenious mechanisms for handling neologisms appropriately.

tion it here in such detail because neologisms are often a stumbling block for translators, as was also evident in the examined sample of translated texts, and Korean very actively creates new ones, often based on words of foreign origin.[74]

In addition to these two essential online explanatory dictionaries, there are other online resources that complement the National Institute of Korean language, namely *Haksŭp Sajŏn* (Korean-Foreign Language Learners' Dictionary).[75] These dictionaries are linked directly to another explanatory dictionary, *Han'gugŏ Kich'o Sajŏn* (Basic Korean Dictionary),[76] which is explanatory, but very limited in scope. In addition to these dictionaries, there is also *Kŭnhyŏndaegugŏ Sajŏn* (Modern and Contemporary Korean Language Dictionary),[77] which contains old expressions with interpretation and offers the possibility of searching for syllables in older spellings, i.e. even with extinct letter combinations. Those interested in dialects will be pleased to see the *Chiyŏgŏ Chonghap Chŏngbo* (Comprehensive regional language information) section on the website of the National Institute of Korean Language, which also contains a dictionary[78] including a map of the distribution of the most common dialect words, etc. Other dictionaries are more relevant for specialists, such as the sign language dictionary. All these dictionaries provided by National Institute of Korean language can be used in a classic way in teaching or research.[79]

In addition to dictionaries, the National Institute of Korean language also builds corpora, which are the basis of modern language research. The *Uri moduŭi malmungch'i* (Modu Corpus, lit. A corpus of all of us)[80] project is tasked with the modern digitization and research of the Korean language. At the time of writing, there are 62 sub-corpora of various kinds on the Modu Corpus site, such as parallel translation corpora[81], tagged

[74] The creation of shortcuts and blends is particularly specific, as described by, for example, Anna Borowiak: A. Borowiak, "On Some Selected Problems of Korean Abbreviology," *Journal of Korean Humanities and Social Sciences*, vol. 1 (2015).

[75] 학습사전, https://krdict.korean.go.kr/. At the time of writing, they were created for four European languages (English, French, Spanish, Russian), Arabian and five Asian languages (Japanese, Thai, Indonesian, Chinese, Mongolian).

[76] 한국어기초사전, https://krdict.korean.go.kr/. The dictionary is providing basic Korean vocabulary of approx. 50,000 items that are essential for learning Korean language (https://www.korean.go.kr/front_eng/activ/activ_03.do, 5 XII 2024).

[77] 근현대국어사전, https://opendict.korean.go.kr/modern/main, 5 XII 2024.

[78] 지역어 종합 정보, https://dialect.korean.go.kr/dialect/, 5 XII 2024.

[79] The tab *Sajinŭ-ro Ponŭn Saenghwarŏ* (Lifestyle Words Seen in Pictures) can be useful for teaching Korean culture, which offers traditional Korean products such as fans, kimchi and brushes and their production process, raw materials, etc. in photos, but only in Korean, so it is only suitable for more advanced students (https://dialect.korean.go.kr/dialect/pub/region_culture, 5 XII 2024).

[80] 우리 모두의 말뭉치, https://www.youtube.com/watch?v=98Bn1vf7HmI, https://kli.korean.go.kr/, moved from former corpus.korean.go.kr, 5 XII 2024.

[81] Parallel corpora of Korean and other languages that are key to machine translation include Russian, Cambodian, Filipino, Thai, Indonesian, Hindi, Uzbek, and Vietnamese (https://kli.korean.go.kr/corpus/introduce/corpusStatistics.do, 5 XII 2024).

corpora, corpora focused on a certain type of text such as newspaper, online communication, messenger texts, etc.[82]

The description of Modu Corpus also mentions AI: *A corpus is a database input to collect language data from various fields, analyze and process the data with a computer, and can be used for lexicography, language education, language research, and artificial intelligence development.*[83] A significant disadvantage is that the corpora are not freely accessible for public search and viewing, not even in a restricted visitor mode.[84]

Finally, let's mention the *Han'gugŏ Haksŭpcha Malmungch'i* (Korean Learner Corpus)[85], a relatively new project of corpora made up of Korean language students' texts, it is intended for the research of specific differences and the analysis of students' errors in a certain language. It is of great importance, for example, in machine translation training. As for other corpus materials, a number of Korean universities offer access to their resources on their websites, or to applications that work with the National Institute of Korean language resources[86] or their own.[87]

The language resources provided by the National Institute of Korean language are very numerous and very well accessible, except for the mentioned difficulties with the use of Modu Corpus. In particular, parallel translation corpora and the Learner Corpus can be understood as a source for the future development of the translation functions of machine translation programs, however, they are currently only available for a small number of languages, and the further development and use of these sources must therefore be monitored.

---

[82] "Modu Corpus offers a variety of raw corpus, including newspapers, books, daily conversations, messenger conversations, and web materials, as well as annotated corpus with a variety of information such as morphology, word sense, syntax, and named entities." (https://kli.korean.go.kr/corpus/boards/faqList.do, 5 XII 2024).

[83] Ibid.

[84] Which is allowed, for example, by the Czech National Corpus (https://www.korpus.cz/, 5 XII 2024), the British National Corpus (http://www.natcorp.ox.ac.uk/, 5 XII 2024) and many others after registration. Regarding the Korean corpora on the website of the National Institute of Korean language, you need to choose a corpus and apply for access, which is allowed for 1 month as a standard (information from 2024). This is somewhat surprising in the context of other freely available digitized materials, as eg. Cha mentions (J. Cha, "The Digital Turn...").

[85] 한국어 학습자 말뭉치, https://kcorpus.korean.go.kr/, 5 XII 2024.

[86] Project at the Soul National University *Kkokkoma* (꼬꼬마) using Sejong Corpora by National Institute of Korean language (http://kkma.snu.ac.kr/); Lee Dongjoo, Jongheum Yeon, Inbeom Hwang, Sang-goo Lee, "KKMA: A Tool for Utilizing Sejong Corpus Based on Relational Database," *Journal of KIISE: Computing Practices and Letters*, vol. 16, no. 11 (2010), pp. 1046-1050.

[87] Yonsei University, Institute for Language and Information studies (언어정보연구원) created *Yŏnse 20Segi Han'gugŏ Malmungch'i* (Yonsei 20th Century Korean Corpus): https://ilis.yonsei.ac.kr/corpus/#/, 5 XII 2024). Unfortunately, one of the early corpora of the Korean language working at the beginning of the new millennium, created at the Korea Advanced Institute of Science & Technology, has not been accessible for some time (http://semanticweb.kaist.ac.kr/home/index.php/KAIST_Corpus, 5 XII 2024).

In addition to the aforementioned National Institute of Korean language, specialized university centers are also involved in the development of various language tools. As an example of very useful tools when working with Korean provided by universities, whether in teaching or in scientific activities, we want to mention the possibilities and limits of the tools developed at the Department of Information and Computer Engineering at Pusan National University. These are concentrated on the *Urimal Paeumt'ŏ* (Korean language learning center) website,[88] where there are directions for the mentioned applications. We will not deal in detail with their dictionaries here but only with the main applications, which are as follows:

- *Han'gugŏ Match'umbŏp/Munbŏp Kŏmsagi* (Korean spelling/grammar checker, later in the text referred to as Speller)[89]
- *Romaja Pyŏnhwan'gi* (Roman character converter, later in the text referred to as Roman)[90]
- *P'yojunbarŭm Pyŏnhwan'gi* (Standard pronunciation converter, later in the text referred to as Pronunciation),[91] and
- *Oeraeŏ – Han'gŭl P'yogi Sangho Pyŏnhwan'gi* (Foreign language words – Hangul notation converter, later in the text referred to as Loanword)[92].

Speller is a Korean text spelling checker that is also used commercially. Roman includes three romanization systems – Revised Romanization, McCune Reischauer and Yale system, and their application in five different semantic or functional groups. Pronunciation converts words, typically containing more complex pronunciations such as words with assimilations, into their phonetic transcription. A secondary function, but very valuable for students and researchers, is the conversion of Korean words to IPA, which is usually quite laborious for students and researchers too. Loanword translates words in foreign languages into Hangul, i.e. it transcribes the original Latin expression into Hangul script and vice versa, i.e. it identifies the Hangul expression in its original form in Latin.

We will now focus on these tools and their functions in terms of their accuracy and reliability when used in study or scientific work. We will present them now in a little more detail.

Speller is a tool for checking the spelling of texts in Korean, so it is a great helper in the process of creating texts in Korean. It identifies incorrect graphemes (e.g. letter swaps such as /e/ and /ae/), letter shifts within syllables when typing quickly, as well as the complicated area of gaps between words and phrases. The correction takes place in the form of suggestions that are displayed above the identified error in a dialog window. Suggested corrections are further explained with links and citations of Korean spelling rules,[93] and

---

[88] 우리말 배움터, http://urimal.cs.pusan.ac.kr/urimal_new/, 5 XII 2024.

[89] 한국어 맞춤법/문법 검사기, http://nara-speller.co.kr/speller/, 5 XII 2024.

[90] 로마자 변환기, http://roman.cs.pusan.ac.kr/, 5 XII 2024.

[91] 표준발음 변환기, http://pronunciation.cs.pusan.ac.kr/, 5 XII 2024.

[92] 외래어 ↔ 한글 표기 상호 변환기, http://loanword.cs.pusan.ac.kr/, 5 XII 2024.

[93] Korean orthography (한국어 맞춤법), https://korean.go.kr/kornorms/regltn/regltnView.do#a, 5 XII 2024.

can be accepted or rejected by the user. However, Speller cannot evaluate an error e.g. inappropriate politeness, i.e. impolite endings in a formal text or in a sequence of sentences that are polite and followed immediately by the impolite Panmal style. So, it is obvious that the multi-sentence context, which is the domain of AI driven tools and helps very well in translation, is not yet fully adapted by Speller. On the other hand, Speller can also in some cases correct or suggest a more appropriate expression within the context of one sentence.

We briefly compare it with a similar tool, Match'umbŏp Kŏmsagi (Spelling checker), offered by Daum, whose translator is included in the machine translation analysis above.[94] However, Daum spell checker was found less accurate or strict compared to Pusan Speller, and its correction suggestions were quite different. It was unable to correctly evaluate errors at the level of wrong letters and shifts of letter in syllables, and in particular, it always evaluated errors in the division of words as two different expressions, regardless of the meaning of the sentence. On the contrary, Speller also provides suggestions for more appropriate vocabulary and thus works with the context of the meaning of the entire sentence, which makes it a much more accurate tool. Overall, Speller excels in both accuracy and soundness over Daum's spell checker, providing more accurate suggestions and also more of them.

Pusan's Romanizator includes conversion into three romanization systems – Revised Romanization, McCune Reischauer and Yale system, and their application in five different semantic or functional categories. There are a large number of different romanization applications available on the internet, whether focused only on Korean or on several languages, between which the script conversion is performed. However, in the case of Korean, they are usually focused on only one romanization system. So Roman stands out among them because it includes the three most used systems. The function of converting to five categories 'personal name', 'proper nouns', 'administrative areas', 'general things' and 'academic application' is completely unique. The last option is advantageous when converting whole sentences, for example, into educational or scientific texts. It also contains a help function and an explanation of how the category is converted. It can also be used in the English version, which works just as well as the Korean one. It is therefore usable in this respect even for users who do not know Hangul, who, for example, come across a Korean name of a place or a name that they need to translate or romanize. However, like most programs, it also warns of a possible error rate, which has been confirmed in some cases.

Application Pronunciation converts words, typically containing more complex pronunciations such as words with assimilations, into their phonetic transcriptions in Hangul and IPA, the overview of which can be opened on the results page. You can enter a word or a whole sentence and the answer contains both a correction in the case of a wrongly entered unit, transcriptions and a link to the relevant rules in the help window. E.g. when inserting a test word containing nasalization, the help links in great

---

[94] Daum Spelling Checker (맞춤법 검사기) can be found as one of the links in the Sajŏn (Dictionary) section when accessing from the main website: https://dic.daum.net/grammar_checker.do, 5 XII 2024.

detail to the relevant article with a number of additional examples. The transcription is in both Hangul and IPA, which is convenient for transcription when working especially on phonetic and phonological topics in undergraduate and other qualification theses, as well as for researchers in this field. Its accuracy was tested on a set of words containing different types of assimilation, just like the previous Roman. Pronunciation achieved better results even for words that Roman romanized incorrectly. However, a big disadvantage compared to Roman is the fact that the transcription is only given in Hangul and IPA, not in any of the transcriptions offered by Roman. It is therefore a tool more suitable for teaching and researching Korean phonetics and phonology by a Koreanist with knowledge of Korean language, as unlike Roman, it does not have an English version.

Pusan Loanwords converts words in foreign languages into Hangul, and converts defined Hangul words into the Latin alphabet. However, just like Pronunciation, it does not have an English version and it is necessary to follow the Korean instructions. It also translates only the words it has in the dictionary and, in the case of a match, offers a selection or similar words containing part of the entered expression. It follows rules that it identifies in detail.[95] When entering a query, it is possible to select the area that the word refers to in order to refine the result.[96] The use of Loanword has significant limitations, unlike the previous applications, which are more universal, because it only works with words that are listed in its dictionary. When entering queries about local names such as Warszawa, Krakow, Praha, it translates flawlessly, it also works when entered in reverse. However, if used for the names it only worked in the cases of some foreign famous people[97], and did not transcribe Korean names into the Latin alphabet.[98] In this regard, its functions are more limited than those of Roman.

The tools presented above can be used with advantage for work with Korean language during study, teaching or research. However, each tool has certain limits and problems, which can be reduced when combined with each other, but at the same time they are a kind of warning against uncritical acceptance of the results that these applications offer us, just like in the case of machine translation mentioned before.

CONCLUSION

In this article, I have tried to introduce some essential AI-powered tools available to Koreanists and those researching Korea and the Korean language. The tools are mainly from the field of programs providing machine translation, of which there are already

---

[95]   https://www.korean.go.kr/front/page/pageView.do?page_id=P000104&mn_id=97, 5 XII 2024.

[96]   Personal name, administrative area, chemistry, general things and All (shows all results): http://loanword.cs.pusan.ac.kr/, 5 XII 2024.

[97]   E.g. Albert Einstein can be converted without any problems, but not Blanka Ferklová.

[98]   While in the categories of personal names, administrative areas, Loanwords converts between Hangul and Latin, in the category of chemistry it translates the given term into or from Korean, which makes it somewhat of a specialized chemistry dictionary.

dozens in the case of Korean. Some of the most common and well-known programs, i.e. Naver Papago, Daum Translate, Google Translate, DeepL Translate, Microsoft Bink, as well as the lesser-known Systran and Promt, including ChatGPT, were tested for known issues that machine translation of Korean to English usually suffers from. First of all, for the recognition and translation of phrasemes already mentioned by Bar-Hillel, a pioneer of machine translation since the 1950s, followed by homonyms and polysemes, personal names, neologisms, Chinese characters, etc. Translators generally had the largest problem with idioms, neologisms, and many of them also with Chinese characters (surprisingly, for example, Korean translator Daum failed completely).

This analysis also focused on other problems which in addition to the machine translators' common difficulties, were caused by the de facto double standard of the Korean language. It tested the machine translators' ability to identify and correctly translate North Korean newspaper texts, which differ from the South Korean standard mainly by spelling differences, also vocabulary and terminology, and to a lesser extent also by morphology. The results showed that the Korean translators like Daum and Papago generally performed worse in this area than machine translators like Google Translate or DeepL Translate. The machine translators were also very sensitive to typesetting, especially gaps in the text, which required careful pre-editing and post-editing of the results, thereby reducing the advantage of the machine translation's speed of translation.

The translators were tested on a small sample of newspaper and scientific texts, and the aim of the analysis was only to illustrate the overall usefulness and applicability of the translation from a reader's point of view, usable by students and researchers working with Korean texts. From this point of view, the Google Translate and DeepL Translate performance was the most flawless. However, despite their great usefulness, it is obvious that these translators cannot be used as the only tool to translate Korean texts.

This brief analysis of several machine translation programs, intended to illustrate its current state, was supplemented by a more detailed description of other tools for working with Korean texts and the Korean language. The tools and resources provided by the National Institute of Korean language were introduced, especially the numerous online dictionaries, corpora and some other resources. At the end of this paper, the tools developed at the Department of Information and Computer Engineering at Pusan National University were presented, which focus mainly on Korean orthography and the mutual relationship between the Korean script hangul and the Latin alphabet. Like every tool mentioned so far, these too have many limitations. E.g. Speller, a spell-checking tool, does not take into account the broader context of a text, Loanword does not romanize or translate into Hangul words that are not in its dictionary, Roman makes errors in romanization of words with exceptional pronunciation, Pronunciation does not translate words that is not in its limited dictionary. Knowing these limits, however, it is very convenient to use these tools and they greatly contribute to the quality of work and save time.

Overall, this article wanted to provide an insight into the state of tools that can be used for studying or researching the Korean language, especially nowadays when the use

of weak AI in working with texts is already common. However, as it follows from the analyses, no matter how advanced translators or tools are, they are not able to automatically translate or evaluate texts flawlessly without the user being aware of their limitations. The following quote is a comment on the content created by GenAI, but it can also be applied to translations that make up machine translation: *While tools like ChatGPT can often generate answers that sound reasonable, they cannot be relied upon to be accurate. Most often, the errors will go unnoticed unless the user has a solid knowledge of the topic.*[99]

REFERENCES

Ahn Miyoung, "MT Problems and Its MTPE Suggestion with Regard to Structural Differences between English and Korean," *The Journal of Mirae English Language and Literature*, vol. 25, no. 1 (2020), pp. 103-130.

Bar-Hillel Y., "Some Linguistic Problems Connected with Machine Translation," *Philosophy of Science*, vol. 20, no. 3 (1953), pp. 217-225.

Bojar O., *Čeština a strojový překlad*, Prague 2017, Studies in Computational and Theoretical Linguistics, 11.

Borowiak A., "On Some Selected Problems of Korean Abbreviology," *Journal of Korean Humanities and Social Sciences*, vol. 1 (2015).

Cha J., "Digital Korean Studies: Recent Advances and New Frontiers," *Digital Library Perspectives*, vol. 34, no. 3 (2018), pp. 227-244.

Cha J., "The Digital Turn in Korean Studies: Trends, Challenges, and Prospects," *Youtube*, 15 IX 2021, at https://youtu.be/sheRZim_wwc?si=ZafNJJVFez8TbK_N.

I Ung, "In'gongjinŭngyong Han'gugŏ malmungch'i 155 Ŏgŏjŏl kuch'uk 5 nyŏn'gan 175Ŏk Chiwŏn" (Establishment of 15.5 Billion Korean Words for Artificial Intelligence... 17.5 Billion Won in Support over 5 Years), *Yonhap News*, 9 IX 2017, at https://www.yna.co.kr/view/AKR20171008048600005.

Jung J., "Top 6 Gen AI Projects from South Korean Tech Giants," *KoreaTech Today*, 31 X 2023, at https://www.koreatechtoday.com/a-glimpse-into-2023-top-5-gen-ai-projects-from-south-korean-tech-giants/.

Kwak Baštanová Y., "Grammatical Expressions of Politeness in Czech and Korean: Tykání and Vykání in Korean," PhD diss., Palacký University Olomouc, 2021.

Kyung Mi, "Samsung Develops In-House AI Platform for Employees, Avoiding Leaks and Enhancing Data Privacy," *KoreaTech Today*, 16 V 2023, at https://www.koreatechtoday.com/samsung-develops-in-house-ai-platform-for-employees-avoiding-leaks-and-enhancing-data-privacy/.

Lee Chang-Soo, "Ch'aetGPT Ch'urhyŏn Ihu Kigye Pŏnyŏkkwa In'gan Pŏn Station Kanŭi Pŏn Station Munch'e Ch'ai Pyŏnhwa Yŏn-gu" (A Follow-up Study of Stylistic Differences between Human and Machine Translation with ChatGPT Added in the Mix), *Beonyeokgyeongu*, vol. 24, no. 3 (2023), pp. 539-561.

---

[99]   Miao Fengchun, W. Holmes, "Guidance...," p. 8.

Lee Dongjoo, Jongheum Yeon, Inbeom Hwang, Sang-goo Lee, "KKMA: A Tool for Utilizing Sejong Corpus Based on Relational Database," *Journal of KIISE: Computing Practices and Letters*, vol. 16, no. 11 (2010), pp. 1046-1050.

Lee Jeong-Hwa, Cha Kyung-Whan, "An Analysis of Korean-English Translation Errors in Google Translate," *The Journal of Linguistic Science*, no. 89 (2019), pp. 221-257.

Lee Noh-shin, Lee Sinjae, Lee Jaeyeong Lee Juhui, "Tongbeonyeogui Miraejipyeong: In-gongjineunggwa Sotonghyeong Yunghaptongbeonyeong Yeongu" (Future Horizons of Translation & Interpretation: Artificial Intelligence and Interactive-Converged Translation & Interpretation), *Beonyeokagyeongu*, vol. 17, no. 2 (2016), pp. 65-89.

Lewis J., "Globalization, Culture and Technology: The Case of Korean Machine Translation," *Hitotsubashi Journal of Social Studies*, vol. 35, no. 1 (2003), pp. 17–28.

Lim Soon Jeung, "Kugŭl shin'gyŏngmang pŏnyŏng ŏlmana chinhwahaenna? Ch'och'anggiwa hyŏnjaeŭi pŏnyŏng p'umjirŭl chungshimŭro" (What Progress Has Been Made in Neural Machine Translation: The Case of Google Translate), *P'ŭrangsŭŏgwŏn munhwayesuryŏn'gu*, vol. 83 (2023), pp. 163-200.

Lim Soon Jeung, Mi-sun Han, "Kugŭl Pŏnyŏgŭi Saryeyŏn-gu: Pŏnyŏkp'umjilgwa Pŏnyŏktan-wiwaŭi Kwan'gyerŭl Chungshimŭ-ro" (A Case Study on "Google Translate: The Correlation between Translation Quality and Units of Translation in Machine Translation), *Beonyeokagyeongu*, vol. 15, no. 1 (2014), pp. 177-209.

Lim Tai Wei, "North Korea's Artificial Intelligence (A.I.) Program," *North Korean Review*, vol. 15, no. 2 (2019), pp. 97–103.

Miao Fengchun, Holmes W., "Guidance for Generative AI in Education and Research," *UNESCO*, at https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research.

Nam Kilim, Lee Soojin, Jung Hae-Yun, "The Korean Neologism Investigation Project: Current Status and Key Issues," *Dictionaries: Journal of the Dictionary Society of North America*, vol. 41, no. 1 (2020), pp. 105–129.

"National Strategy for Artificial Intelligence," *Ministry of Science and ICT*, at https://www.msit.go.kr/bbs/view.do?sCode=eng&nttSeqNo=9&bbsSeqNo=46&mId=10&mPid=9.

Park Junsuk, "A Comparative Study of AI translation – Google Translate and Papago," *Studies in Linguistics*, vol. 65 (2022), pp. 93-111.

Park Junsuk, "Analysis of the Types of Translation Errors by the Google Translator," *The Jungang Journal of English Language and Literature*, vol. 59, no. 4 (2017), pp. 279-299.

Park Minseo, "Zuckerberg's South Korea Visit: Collaborating on AI and XR Technologies with LG and Samsung," *KoreaTech Today*, 28 II 2024, at https://www.koreatechtoday.com/zuckerbergs-south-korea-visit-collaborating-on-ai-and-xr-technologies-with-lg-and-samsung/.

Park Soojung, Eunsil Choi, "Ch'aetGPTŬi Airŏni Pŏnyŏk Hwaryong Kanŭngsŏng Koch'al" (A Study of Translatability of Irony in ChatGPT), *Beonyeokagyeongu*, vol. 24, no. 2 (2023), pp. 131-160.

Pimienta D., Prado D., Bianco A., „Twelve Years of Measuring Linguistic Diversity in the Internet: Balance and Perspectives," *UNESCO*, at https://unesdoc.unesco.org/ark:/48223/pf0000187016.

Pucek V., *Gramatika korejského jazyka*, Praha 2005.

Pym A., "AI and the Future of Translation: In Search of Evidence," *Youtube*, 18 VII 2023, at https://youtu.be/46rUOOmNlkA?si=s8uGcDkIneM5fBtd.

Pym A., *Exploring Translation Theories*, London–New York 2023.

Pym A., "GPT in the Training of Translators," *Youtube*, 27 IV 2023, at https://www.youtube.com/watch?v=b9U_FUaneso.

"Q/A," *Choson Ilbo*, at https://news.chosun.com/english/faq.html#:~:text=should%20be%20followed.-,7.,business%2C%20politics%20and%20national%20news

Searle J.R., "Minds, Brains and Programs," *Behavioral and Brain Sciences*, vol. 3, no. 3 (1980), pp. 417-457.

Zeman D., *The World of Tokens, Tags and Trees*, Prague 2018, Studies in Computational and Theoretical Linguistics.

Newspaper articles and texts translated and analyzed

Hŏ Jiwŏn, "Minjudang 'Pulbŏpchuch'a Ch'ŏbŏl Nanimch'iryo Yugŭp Hyu-ga Hwaktae' tŭng Kongyak" (Democratic Party Pledges to Punish Illegal Parking and Expand Paid Leave for Infertility Treatment), *Naver News*, 4 II 2024, at https://n.news.naver.com/mnews/article/079/0003880864.

Kang Hyosim, "Widaehan Ŏmŏniŭi Sarangŭl Chŏtchulgi-ro Hayŏ Chogugŭi Mirae-ga Murŏngmurŏk Charanda" (With the Love of a Great Mother as the Source of Milk, the Future of Our Country Grows Prosperously), *Rodong Sinmun*, 4 II 2024, at http://www.rodong.rep.kp/ko/index.php?QEBAQEAx.

"Kyŏngaehanŭn Kimjŏngŭn dongjikkesŏ Kimhwa-gun chibang gongŏp kongjangdŭrŭl hyŏnji jidohashiyŏtta" (Respected Comrade Kim Jong-un Provided On-site Guidance to Local Industrial Factories in Kimhwagun), *Rodong Sinmun*, 4 II 2024, at http://www.rodong.rep.kp/ko/index.php?QEBAQEAx.

Park Donggeun, "*Urimal Imojŏmo*" (Everything You Wanted to Know about the Korean Language), *National Institute of Korean Language*, **at** https://www.korean.go.kr/front/etcData/etcDataView.do?mn_id=46&etc_seq=183&pageIndex=14.

Sŏng Hanyonh, "Munjenŭn Ŏgongida...chaembŏri-ga sanŭ-ro kan iyu" (The Problem Are Incompetent Civil Servants – Why Jamboree Why Did Go Wrong), *Hangyoreh*, 20 VIII 2023, at https://www.hani.co.kr/arti/politics/politics_general/1104943.html.

Blanka FERKLOVÁ – an Assistant Professor at the Faculty of Art, Palacky University in Olomouc. She graduated from Charles University in Prague with a degree in Korean Studies. She received a Master´s Degree in Korean studies, Czech Language and Literature and Pedagogy, and her PhD in Asian Languages, specializing in Korean. Her research interests include the modern Korean language, especially lexicology and word formation, with a special focus on Korean ideophones and iconicity in the Korean language. Other research interests are new approaches to the acquisition of the Korean language and its translation.