

Working Group on the Legal Development of the Hague System for the International Registration of Industrial Designs

Eleventh Session
Geneva, December 12 to 14, 2022

STUDY OF THE AVAILABILITY AND USABILITY OF TRANSLATION TECHNOLOGIES

Document prepared by the International Bureau

I. BACKGROUND

1. At its tenth session, held in 2021, the Working Group on the Legal Development of the Hague System for the International Registration of Industrial Designs (hereinafter referred to as the “Working Group”) considered documents H/LD/WG/10/4 and H/LD/WG/10/5, entitled “Update on Advanced Study of the Cost Implications and Technical Feasibility of the Introduction of the Chinese and Russian Languages into the Hague System” and “Update on Criteria for the Selection of Additional Languages for Introduction into the Hague System”, respectively. The Working Group requested the Secretariat to provide a revised advanced study of the cost implications and technical feasibility of the introduction of new languages into the Hague System, taking into account technical solutions, in particular machine translation, for its consideration by the Working Group at its next session (see document H/LD/WG/10/6, paragraph 18(ii)).

2. This document provides detailed information on the translation technologies and practices used by the different translation services at WIPO. The information provided aims to facilitate the considerations of the Working Group in relation to the potential mitigating effects of translation technologies on the costs associated with the introduction of additional languages into the Hague System.

II. TRANSLATION TECHNOLOGIES USED AT WIPO

3. Translation work is performed at WIPO by the different global services for protecting intellectual property as well as the Language Division. The translation process generally consists of a two-step process: firstly the text that is to be translated is matched against entries contained in the applicable translation memory, and secondly a human translator either post-edits translation proposals that are generated by a neural machine translation (NMT) model (WIPO Translate) if there is a model available in that domain and language combination, or translates the remaining text into the target language. Finally, a certain percentage of the translations are subject to quality control by an in-house translator reviser.

WORLDSERVER

4. WorldServer is a translation management system, including a computer assisted translation tool, which manages the translation workflow by automating the translation tasks. It populates firstly the translations with segments in the target language retrieved from the translation memories, and secondly the remaining segments with NMT proposals. Both the Translation and Terminology Section of the Madrid Registry and the PCT Translation Division use WorldServer¹.

TRANSLATION MEMORIES

5. The backbone of each translation practice is the translation memories, which are repositories of previously translated text in the particular domain (Hague, Madrid, PCT, or Language Division). All entries in the translation memories have gone through human validation to ensure quality. The translation memories are kept separate for each service and each language combination². The volume of entries contained in the respective translation memories varies depending on the amount of previously translated text. For example, since 2016, the share of Hague applications filed in English has been over 86 per cent, French around 12 per cent and Spanish around one per cent³. Consequently, the translation memories for Hague translations where English is the source language have much more entries than the translation memories where French is the source language, while the translation memories where Spanish is the source language almost have no entries. The translation memories are integrated into the translation management system (WorldServer).

WIPO TRANSLATE

6. WIPO Translate is an AI-based machine translation tool developed in-house by WIPO and trained for specific subject matter, which uses the latest NMT technology. NMT is a technology which is based on neural network models that “learn” from large volumes of previously translated text⁴. WIPO Translate has initially been trained by and focused on patent documents⁵ in the 10 PCT publication languages⁶. WIPO Translate, which has been made publicly available for users on Patentscope, provides translation proposals from these languages into English and *vice versa*.

¹ The Language Division will switch to WorldServer in the near future.

² For example, there are currently six separate translation memories for the three working languages in the Hague System (English-French, English-Spanish, French-English, French-Spanish, Spanish-English, and Spanish-French).

³ See *Hague Yearly Review 2022*, graph 13.

⁴ For example, the training datasets used by WIPO Translate for the translation of patent documents typically have millions of documents.

⁵ WIPO Translate is trained with text contained in PCT applications and patent collections from national and regional Offices available on Patentscope.

⁶ Arabic, Chinese, English, French, German, Japanese, Korean, Portuguese, Russian and Spanish.

7. A customized model of WIPO Translate has been developed for each of WIPO's different translation services⁷. However, in order to produce high-quality translations, WIPO Translate has to be trained with large volumes of text in the relevant domain for a particular language combination⁸.

8. Furthermore, the translation proposals that are generated by an NMT model for linguistically more distant language combinations are typically of lower quality than for linguistically closer language combinations. Therefore, for distant language combinations, it is even more important to have sufficient training data in order to ensure sufficient translation quality⁹.

TRANSLATION, POST-EDITING AND QUALITY CONTROL

9. The tools described above cannot fully automate the translation process and so, even if machine translation technologies are used, human resources are always required for post-editing and quality control. These translators need to have the appropriate language and translation skills in the respective languages. In that respect, it is recalled that it is always preferable that only a translator whose first language is the target language should perform post-editing and quality control of translations into that language. The quality control is carried out by an internal translator-reviser.

10. The quality of the NMT output has an impact on the costs for post-editing. If the quality of the machine-translation output is poor, the end-to-end costs for post-editing may be close to or even exceed the costs for direct human translation.

III. MACHINE TRANSLATION PRACTICES AT WIPO

HAGUE SYSTEM

11. It is recalled that the translation of international applications under the Hague System is currently performed by the Translation and Terminology Section of the Madrid Registry¹⁰. The translation transactions in relation to Hague applications concern the indication of the product, a description of the characteristic features of the industrial design, claims and legends¹¹. While the product indications, claims and legends usually contain less than 10 words, a description is composed of free text, with an average of 64 words in 2022¹². It is observed that the number of international applications containing a description increased from 17 per cent in 2012 to 65 per cent in 2022¹³.

⁷ A customized model is currently used for Hague translations, Madrid translations, some language combinations in the PCT and for certain documents by the Language Division.

⁸ While WIPO Translate can potentially be trained for one translation direction only, it is usually trained for both directions, as the training with parallel data (two languages of the same content) greatly improves the translation quality.

⁹ For example, English and French are linguistically similar languages, whereas English and Arabic, Chinese or Korean are distant languages. English and Russian are both part of the Indo-European language family; however, they have grammatical differences, which is why more training data would be required for the translation of free text.

¹⁰ All translations (post-editing) of Hague texts are currently carried out in-house.

¹¹ See Article 5(2)(b)(ii) and (iii) of the Geneva Act (1999), Rules 7(3)(iv), (4)(b) and (5)(a), and 11(2) and (3) of the Common Regulations Under the 1999 Act and the 1960 Act of the Hague Agreement (hereinafter referred to as the "Common Regulations"), Sections 403 and 405(c) of the Administrative Instructions for the Application of the Hague Agreement.

¹² The length of a description varies greatly. For example, in 2021 as well as in the first three-quarters of 2022, 15 per cent of the descriptions contained more than 100 words, while one per cent contained more than 500 words. In 2021, 589 Hague applications contained a description which was between 100 and 500 words long, while 30 Hague applications contained a description which was more than 500 words long.

¹³ This development is a result of the accession of Contracting Parties whose Offices are Examining Offices and which require or recommend the inclusion of a description. In this regard, it is noted that China made a declaration requiring that an international application contains a brief description of the characteristic features of the designs, and

12. The Translation and Terminology Section, using a direct translation practice, relies on six Hague translation memories for the three working languages of the Hague System as well as WIPO Translate. The translation memories of previously translated text matter in Hague applications currently hold a total of around 280,000 entries, while the majority is contained in the translation memories where English is the source language. The International Bureau estimates that, on average, for translations from English into French or Spanish, a match is found in the translation memories in around 35 per cent of the text segments, while the remaining text (65 per cent) requires post-editing of the translation proposals generated by WIPO Translate. In comparison, for translations where French or Spanish are the source language, a match is found in the translation memories in much fewer cases¹⁴. Therefore, for such translations, the majority of the text is post-edited using the limited translation proposals generated by WIPO Translate.

13. In addition, it is observed that while the translation memories grow continuously, the matching rate grows much slower. The International Bureau estimates that for Hague text matter, the matching rate in the Hague translation memories may not go beyond 40 per cent, given the continuous development in the field of industrial designs and the related text contained in Hague applications (in particular the free-text descriptions)¹⁵. Therefore, the International Bureau assumes that around 60 per cent of the translations will always be subject to human post-editing of translation proposals generated by WIPO Translate.

MADRID SYSTEM

14. It is recalled that the Madrid System has the same language regime as the Hague System, featuring the same three languages. The majority of translation work performed in the Madrid System concerns the translation of indications of goods and services in international applications. Those goods and services are commonly a list of standardized terms, and therefore differ from the descriptions contained in Hague applications.

15. The Translation and Terminology Section of the Madrid Registry relies on the translation memories of previously translated Madrid text¹⁶ and WIPO Translate. The Madrid translation memories contain more than six million entries in total in the six Madrid translation memories. The matching rate for the translation of Madrid text in the Madrid translation memories is around 65 per cent¹⁷, while the remaining text requires post-editing of the translation proposals generated by WIPO Translate. The quality output of the translation proposals generated by WIPO Translate, in particular for translations where English is the source language, enables the International Bureau to reduce the post-editing costs by 50 per cent¹⁸.

16. Some possible measures are currently being explored by the Working Group on the Legal Development of the Madrid System for the International Registration of Marks (hereinafter the "Madrid Working Group") as part of the discussions relating to a possible expansion of the language regime, such as the possibility of creating translation memories with a large number of equivalent terms in Arabic, Chinese and Russian, training WIPO Translate so that it can handle Madrid translations in those languages, and collecting language pairs of lists of goods and services which Offices of Contracting Parties may have in their possession¹⁹.

the Office of the Russian Federation (ROSPATENT) strongly recommends that a brief description of the characteristic features of the industrial design be provided.

¹⁴ A match is usually only found where the same applicant files multiple applications for very similar designs. In those cases, a match can be found for product indications or legends, but less so for descriptions.

¹⁵ The reasons for this development lie in the nature of the protection of industrial designs which requires novelty and originality under most national laws. Design registration results from new products being developed resulting in new designs, and new designs being created for both emerging and existing products.

¹⁶ Since the Madrid System has the same three working languages as the Hague System, the Madrid Translation and Terminology Section uses six translation memories for the same language combinations.

¹⁷ See document MM/LD/WG/19/7.

¹⁸ This means that the costs for post-editing are 50 per cent of the costs if the text was fully translated without the assistance of WIPO Translate.

¹⁹ See document MM/LD/WG/19/7, paragraphs 39 to 50.

17. However, the different type of text matter contained in Hague and Madrid applications (free text descriptions versus standardized terms), the continuous development in the field of industrial designs, the different volume of entries in the Hague and Madrid translation memories (280,000 entries versus over 6 million entries), and consequently the different matching rates in the Hague and Madrid translation memories as well as the different volumes of potential training data for each WIPO Translate model would make the exploration of similar measures less promising in the context of the Hague System.

PCT SYSTEM

18. The PCT Translation Division is responsible for the translation of titles and abstracts of PCT applications from the 10 PCT publication languages into English and French (if they are not in those two languages), and international search reports and international preliminary reports on patentability into English (if they are not in English)²⁰.

19. The title of an invention is a short, precise description of the invention and is recommended to be between two to seven words long²¹. The abstract of a PCT application, which ranges between 50 to 150 words²², contains a summary of the disclosure of the invention²³. International search reports and international preliminary reports on patentability are technical documents established by patent Offices relating to the patentability of inventions, and, on average, range from 50 to 200 words for international search reports, and from 550 to 950 words for international preliminary reports on patentability, depending on the language in which the report is established. The text matter contained in those reports differ greatly from that contained in Hague applications.

20. The PCT Translation Division relies on the PCT translation memories in the 18 language combinations of the 10 PCT publication languages²⁴, a Terminology Database (“WIPO Pearl”)²⁵ and WIPO Translate. As a result of the differences in volume of PCT applications filed in the respective languages over the years²⁶, it follows that a large amount of entries are available in the translation memories for the language combinations of Chinese, German, Japanese and Korean into English, while in comparison, not many entries are available in the language combinations of Arabic or Russian into English²⁷. Consequently, translations from Arabic and Russian into English are carried out by human translators (with the use of the very limited existing translation memories and the Terminology Database).

21. Likewise, as a result of the different volumes of available datasets for training, WIPO Translate is currently only used in-house for translations from Chinese into English and English into French. The usability of WIPO Translate is currently being assessed for other language combinations (in particular for French, German, Japanese and Korean into English).

²⁰ See Rules 45.1, 48.3(c), 62*bis*.1, 72.1 and 86.2 of the Regulations under the Patent Cooperation Treaty (hereinafter referred to as “PCT Regulations”).

²¹ See Rule 4.3 of the PCT Regulations.

²² If the abstract is in English or when it is translated into English. See Rule 8 of the PCT Regulations.

²³ See Rule 8.1(a)(i) of the PCT Regulations.

²⁴ However, since the PCT Translation Division translates into English, almost no entries are available in the other translation directions, meaning from English into other PCT publication languages (except English-French).

²⁵ The PCT Terminology Database contains scientific and technical terms derived from patent documents in the 10 PCT publication languages. It currently contains around 230,000 terms. It is publicly available at: <https://www.wipo.int/reference/en/wipopearl/>.

²⁶ See *PCT Yearly Review 2022*, graph C1.

²⁷ In addition, international search reports and international preliminary reports on patentability established in Arabic or Russian are not available in machine-readable format, which is why the translations of those documents cannot be included in the PCT translation memories or be used to train WIPO Translate. Therefore, the available data in the PCT translation memories for Arabic and Russian only stem from translated titles and abstracts in PCT applications filed in those languages.

22. The WIPO Translate model customized for internal PCT translations would not be of use for the translation of Hague text matter because, firstly, it is currently only used for two language combinations and, secondly the text matter it was trained with stems from a different technical domain (patent documents).

LANGUAGE DIVISION

23. The WIPO Language Division provides translations from and into the six official United Nations (UN) languages with regard to treaties and other international instruments administered by WIPO and UPOV²⁸, national legislative texts, documents for the Assemblies of Member States, other main bodies, committees and working groups, including their working documents and meeting reports; publications, training material, the WIPO website and other communication content. The text matter translated by the Language Division therefore differs from the text matter contained in Hague applications²⁹.

24. Most translations³⁰ of the WIPO Language Division are carried out from English into the other official UN languages, while a few translations are carried out from French and Spanish into English, and a very small portion of text is translated from the other official UN languages as well as German and Portuguese etc. into English, and *vice versa*³¹. All translations are carried out by human translators, with the support of the respective translation memories. In this respect, it is noted that the translation memories used by the Language Division have data from English into Arabic, Chinese and Russian, however, no data are readily available from any of those languages into English. Since a couple of years, post-editing of machine translation output generated by WIPO Translate has been applied to the translation of certain categories of documents, e.g., verbatim reports.

25. The WIPO Translate model customized for internal use by the Language Division would not be of use for the translation of Hague text matter because, firstly, the text matter it was trained with stems from a different technical domain (meeting documents) and, secondly, the bulk of the translations is carried out from English into other languages which is why no data are available from other languages into English.

IV. OTHER CONSIDERATIONS

QUALITY OF TRANSLATION

26. It is important to recall that the translation of text matter contained in a Hague application is established by the International Bureau without any input or control by the applicant³². In particular, the product indication and the description, or their translation, are then taken into account by certain designated Offices when carrying out substantive examination of the design, and remain relevant for administrative or judicial proceedings in all designated Contracting Parties in respect of refusals, invalidation or infringement. Therefore, the translations established by the International Bureau constitute an important element for any substantive decision regarding the validity or scope of protection in the jurisdictions of Contracting Parties³³, while, unlike in the PCT³⁴, the applicant has no involvement in them.

²⁸ Translation and interpretation services are provided to UPOV as well, which uses four official languages. i.e., English, French, German, and Spanish.

²⁹ While the Language Division is responsible for the translation of Hague related text matters, such as Working Group documents, the text in those documents differs from the text in Hague applications.

³⁰ The WIPO Language Division estimates it to be around 90 per cent.

³¹ Translations from and into German and Portuguese are outsourced.

³² See Rule 6(4) of the Common Regulations. In this respect, it is noted that, while the applicant has the possibility to submit with the international application a proposed translation of any text matter contained in the international application, in practice, the International Bureau never receives such translations.

³³ Contracting Parties may rely on the description as filed or as translated.

³⁴ The abstract of a PCT application, which is translated by the International Bureau, merely serves as a scanning tool for the purposes of searching in the particular art, and is not taken into account for any other purpose,

Hence a high quality of the translation by the International Bureau is imperative for the safeguard of the holders' rights in the territories of designated Contracting Parties.

ASSESSMENT OF NMT TOOLS

27. In addition, the International Bureau carried out a test assessment of the usability of NMT tools, including WIPO Translate which is publicly available on Patentscope, for the translation of text matter contained in Hague applications. The test assessment was carried out for the translation of text matter contained in national Chinese design registrations, which were extracted from the Global Design Database³⁵. The in-house translators' preliminary findings were that the quality output was not sufficient to be used for post-editing, unless the technologies were trained with domain specific text matter. WIPO Translate did, however, show some advantages over other NMT tools when translating technical terms.

28. A further assessment of the usability of WIPO Translate for the translation of Hague text matter in Chinese and other languages would be possible. To that end, however, the International Bureau would need samples of design registrations in additional languages, covering a broad variety of product indications, legends and descriptions³⁶.

MACHINE TRANSLATION SERVICES PROVIDED BY EXTERNAL PROVIDERS

29. Another possibility, if any, would be to rely on machine-translation services provided by external providers. However, the quality of the translation output may not be as high as when using WIPO Translate, given that the quality output depends on the training with a large volume of data in the relevant domain. In this respect, it is recalled that the WIPO Translate model for Hague translations is trained with the corpus of the Hague translation memories. In addition, the use of external machine translation services may also be more costly than WIPO Translate which is developed in-house. Finally, due to the confidential nature of Hague applications, the possible providers would need to be assessed in terms of security and data protection.

V. USABILITY OF MACHINE TRANSLATION TOOLS FOR THE INTRODUCTION OF ADDITIONAL LANGUAGES

30. Currently, the International Bureau does not have any translation memories for the translation of text matter contained in Hague applications in any languages other than the current three working languages. In addition, WIPO Translate has not been trained for Hague terminology in additional languages. Following the analysis of the availability and usability of translation technologies above, the translation of Hague text matter from and/or into any additional languages would be dependent on human translators for the foreseeable future.

31. Furthermore, the usability of these tools for additional languages would depend on several factors, in particular the number of applications filed in that language (which determines the volume of entries in the translation memories and the training data for WIPO Translate), the available volume of other relevant training data in each language combination for the training of WIPO Translate, and the relative similarity/distance of each language combination.

particularly not for the scope of the protection sought (Article 3(3) of the PCT and Rule 8.3 of the PCT Regulations). The scope of protection is defined by the claims, while the description discloses the invention. The claims and descriptions, however, are translated by the applicant upon entry into the national phase before an Office, and the substantive examination by a national/regional Office is carried out based on that translation.

³⁵ The test assessment was carried out for the translation of text matter from Chinese into English because WIPO Translate is used for translations in this language combination in the PCT.

³⁶ In this respect, it is noted that the Global Design Database does not contain any collections in Arabic or Russian.

CONSIDERATIONS RELATING TO THE BUILDING UP OF TRANSLATION MEMORIES

32. If additional languages were to be introduced into the Hague System, the International Bureau would have to establish and build up translation memories for each new language combination.

33. In this regard, building up a corpus of translation memories before the introduction of an additional language may be a means to enhance their usability and could reduce the translation costs for languages if a sufficient number of filings can be forecasted in those languages. The upfront translation of the current entries in the translation memories into additional languages, as is currently under consideration by the Madrid Working Group for the indications of goods and services, could therefore be a possible option. However, while the creation of equivalent terms in additional languages may be of use for shorter text segments (in particular product indications or legends), it would not be of the same assistance for the translation of descriptions³⁷, which account for the bulk of the translation work.

34. Furthermore, the translation of the entries in the translation memories would come at cost³⁸. Therefore, the cost of translating the existing translation memories into additional languages would need to be compared with the cost of simply translating text matter in Hague applications in additional languages and building up translation memories over time. Below a certain level of filings, the cost of translating the existing translation memories would be significantly higher than translating text matter in Hague applications filed in additional languages.

TRAINING OF WIPO TRANSLATE

35. The International Bureau currently relies on WIPO Translate for the machine-translation of Hague text matter that is not covered by the translation memories. The use of WIPO Translate for new language combinations would be an available option.

36. In this regard, in order to gain sufficient quality for the translation of Hague text matter, WIPO Translate would need to be trained with large volumes of text in the relevant domain for each new language combination. Thus, the sharing of any relevant language combinations of terms such as product indications, legends, claims, and descriptions would be of assistance. It is however uncertain as to whether large volumes of such training data are in the possession of the Offices of Contracting Parties themselves or available elsewhere. It is also to be underscored that while WIPO Translate can be trained with entries in one translation direction only (for example, English into Chinese), this does not guarantee the same quality of translation if it is used for translations in the other translation direction (for example, Chinese into English)³⁹. Moreover, the upfront translation of relevant text matter would come at cost in any case, if it were to be done by the International Bureau⁴⁰.

37. Furthermore, it is recalled that, despite considerable volumes of potential training data being available, the PCT System currently uses WIPO Translate for two language combinations only⁴¹, while the Language Division currently only uses WIPO Translate to a very limited extent⁴². Therefore, at this stage, the potential usability of WIPO Translate for the translation of Hague text matter beyond the current three languages appears to be quite limited.

³⁷ As mentioned above, descriptions consist of free-text and differ greatly. See also paragraph 17.

³⁸ The costs for the translation of the Madrid translation memories into Arabic, Chinese and Russian are estimated to be between 2.36 million Swiss francs and 4.31 million Swiss francs. See MMLD/WG/19/7, paragraph 42.

³⁹ See footnote 8.

⁴⁰ See paragraph 34.

⁴¹ See paragraph 21.

⁴² Certain categories of documents, such as verbatim reports of WIPO meetings. See paragraph 24.

38. Finally, it is also noted that NMT, including WIPO Translate, is advancing, and a periodical assessment of the available NMT tools for the translation of Hague text matter could therefore be recommended.

VI. IMPLICATIONS ON COST ESTIMATES

39. The recurrent costs for the translation of Hague text matter estimated in document H/LD/WG/10/4 were based on the rates applicable in the PCT for the outsourcing of translations⁴³. The rates are based on the volume of the translated text per language combination, the technical nature of the translated text and the specifics of each language combination in the PCT, and therefore the actual rate that would be applicable for each of the additional languages would be subject to change in the context of the overall situation of the Hague System, if they were to be introduced. In this regard, the present analysis focuses on the potential mitigating effects of translation technologies currently available on such recurrent translation costs associated with additional languages.

40. The cost estimates in that document already took into account the efficiencies gained through the gradual compilation of translation memories. Even if the International Bureau translated the corpus of the translation memories into additional languages, this is not likely to increase the matching rate for the translation of descriptions, and the upfront translation would come at a cost⁴⁴. In fact, given the continuous development in the field of industrial designs and the related text contained in Hague applications, the International Bureau observes a decelerated growth of the matching rate in the Hague translation memories, and therefore estimates that around 60 per cent will always be subject to human post-editing of translation proposals generated by an NMT tool⁴⁵.

41. Moreover, in view of the present analysis, the currently available NMT tools would not, in the foreseeable future, improve the productivity or efficiency of translations for text matter contained in Hague applications in additional languages. The cost of translations and quality control could not be reduced without an impact on the quality of the translation. In this respect, it is emphasized that any compromise on the quality of the translation would jeopardize the rights of users⁴⁶. Therefore, the technical analysis does not lead to any revised cost estimates for the introduction of additional languages into the Hague System.

42. Finally, it is noted that the use of translation technologies can only have implications on the cost of the actual translations. All other costs, such as the cost of servers to support any additional translation workload and quality control, would occur regardless of the use of translation technologies if additional languages were introduced into the Hague System.

43. The Working Group is invited to discuss and comment on the content presented in this document.

[End of document]

⁴³ The cost estimate showed the additional translation cost for the introduction of Chinese and Russian, as requested by the Working Group (see document H/LD/WG/8/8, paragraph 29). While it can be assumed that the same recurrent translation costs of introducing other languages could also be based on the rates applicable in the PCT, a further assessment could be carried out, if so endorsed by the Working Group.

⁴⁴ See paragraphs 34 and 36, above.

⁴⁵ See paragraph 13, above.

⁴⁶ See paragraph 26, above.