MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230933

Translation of Ontological Concepts from English into German Using Commercial Translation Software and Expert Evaluation

Richard NOLL[a,1], Alexandra BERGER[b], Carlo FACCHINELLO[c], Oya GÜNGÖZE[c,d], Michael von WAGNER[c,d], Sebastian HOEHL[e], Michaela NEFF[a], Holger STORF[a], Jannik SCHAAF[a]

[a] *Goethe University Frankfurt, University Hospital Frankfurt, Institute of Medical Informatics, Frankfurt, Germany*
[b] *Frankfurt Reference Centre for Rare Diseases, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany*
[c] *Department of Internal Medicine 1, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany*
[d] *Department of Medical Information Systems and Digitalization, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany*
[e] *Institute of Medical Virology, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany*
ORCiD ID: Richard NOLL https://orcid.org/0000-0002-5240-8484

Abstract. Medical ontologies are mostly available in English. This presents a language barrier that is a limitation in research and automated processing of patient data. The manual translation of ontologies is complex and time-consuming. However, there are commercial translation tools that have shown promising results in the field of medical terminology translation. The aim of this study is to translate selected terms of the Human Phenotype Ontology (HPO) from English into German using commercial translators. Six medical experts evaluated the translation candidates in an iterative process. The results show commercial translators, with DeepL in the lead, provide translations that are positively evaluated by experts. With a broader study scope and additional optimization techniques, commercial translators could support and facilitate the process of translating medical ontologies.

Keywords. Medical ontology, translations, controlled vocabulary

1. Introduction

Ontologies in medicine pursue the consolidation of knowledge as an accessible entity that enables the standardization of information. Ontologies are used for language and semantic links between terminologies and describe a kind of network of relationships. The standardized form with its hierarchical structure enables all professionals in the given field to align the desired outcomes in medical practice [1].

Currently, commonly used ontologies, such as SNOMED Clinical Terms (CT) or those included in the Unified Medical Language System (UMLS) terminology database,

[1]Corresponding Author: Richard Noll, Institute of Medical Informatics, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; richard.noll@kgu.de.

are mostly available in English and a few other languages [2, 3, 4]. This presents a language barrier that is a limitation in research and automated processing of patient data.

Manual translation of an already existing ontology is very complex and time-consuming [2]. Machine translation reduces the effort by providing an automated solution to quickly and accurately translate large amounts of text. Algorithms can analyze text and identify patterns to generate an accurate translation. Freely available commercial translation tools (e.g., Google or DeepL) have shown promising results in previous studies in the field of medical terminology translation [3, 4]. The studies already conducted are limited to only a few translators and mostly refer to SNOMED CT.

This study is intended to provide a proof-of-concept and to demonstrate the feasibility of representing English-language ontologies in German using commercial translators. In this study, Human Phenotype Ontology (HPO) is used as the base ontology. HPO offers comprehensive bioinformatics resources for the analysis of human diseases and phenotypes. It has become a standard used by several clinical genetics' consortia [5].

2. Methods

2.1. Study design

The purpose of this study is to investigate the outcome of machine translation of English ontology terms and synonyms into German. Several pipelines with a total of 12 different commercial translation APIs were set up via Python. In the procedure shown in Fig. 1, 130 HPO terms and associated synonyms were translated and evaluated. The evaluation was performed by 6 medical experts from the University Hospital Frankfurt, based on a 4-point Likert scale with the possibility to provide additional comments. The experts' assessments are ranked and discussed by the author of this paper. The study started in July 2022 and was conducted over a four-month period ending in November.

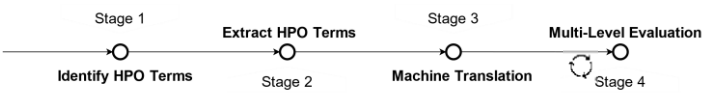

Figure 1. Iterative four-stage process with (1) identifying English HPO terms, (2) extracting these HPO terms, (3) translating terms into German using machine translation and (4) evaluating the translations. The evaluation is carried out in a multi-stage and independent curation process (Multi-Level Evaluation).

2.2. Human Phenotype Ontology

The HPO provides a standardized vocabulary for phenotypic abnormalities that occur in human disease. Each term in the HPO refers to a phenotypic abnormality and has a specific structure, see Fig. 2. HPO currently includes over 13,000 terms. There are over 123,000 annotations of terms for rare diseases and over 132,000 for common diseases [6]. The HPO is freely available and can be downloaded from the HPO website [7].

id: HP:0000091
name: Abnormal renal tubule morphology
def: "An abnormality of the renal tubules."
comment: The renal tubules are reabsorptive canals that are involved in the secreting, collecting, and conducting of the urine.
synonym: "Abnormality of the renal tubule"
synonym: "Morphologic abnormality of the renal tubules"
xref: UMLS:C4021826
is_a: HP:0012575 ! Abnormal nephron morphology

Figure 2. Structure of the HPO using the term "HP:0000091". Additional annotations can be definitions, comments, synonyms, cross-references and ontology-typical hierarchical "is_a" dependencies [7].

2.3. Implementation

Python 3.9.13 is used for the implementation. A tool (*term randomizer*) was written that randomly selects any given number of HPO terms. At first, 10 randomly selected terms were translated. After the first round of expert evaluation, a pipeline was created with the five translators who had achieved the best ratings. A list of 100 random and 20 frequent terms was created and translated by these top five translators. The 20 frequent terms were identified by medical professionals out of 178 physicians' letters from the Frankfurt Reference Centre for Rare Diseases at the University Hospital Frankfurt.

Various translators were used for the translation from English into German, see Table 1. Except for DeepL, all translators were accessed through the *translators* library (version 5.4.1) [8]. All translators in the library that support translation from English into German and whose API was accessible at the time of the study were used. For DeepL, a direct link to their API was implemented [9].

The baidu translator allowed the specification of the translation in terms of a medical focus (*professional_field* = 'medicine'). This enables the translation to match those derived from a corpus of documents within the medical field.

Table 1. Different translators used in this study, their origins, the underlying companies, and their websites [8].

Translators	Origin	Company	Website
Argos	USA	P.J. Finlay	translate.argosopentech.com
Baidu	China	Baidu	fanyi.baidu.com
Bing	USA	Microsoft	bing.com/translator
DeepL	Germany	DeepL	Deepl.com
Google	USA	Google	translate.google.com
Iciba	China	Kingsoft/ Xiaomi	iciba.com/fy
Iflytek	China	Iflytek	fanyi.xfyun.cn/console/trans/text
Itranslate	Austria	Itranslate	itranslate.com
Reverso	France	Reverso	reverso.com
Tencent	China	Tencent	fanyi.qq.com
Sogou	China	Sogou / Tencent	fanyi.sogou.com/text
TranslateCom (TC)	USA	TranslateCom	translate.com

2.4. Evaluation

The evaluation of the translations was performed in a two-step evaluation process carried out by six different medical experts. The evaluation of the translations was done via a 4-point Likert scale: strongly agree = 1, agree = 2, disagree = 3, strongly disagree = 4. The scale describes the approval of the expert for the respective translation. In addition, the medical experts could make comments on individual term or synonym translations. The comments made were aggregated and categorized in the results section.

For the first evaluation step, the translations of the 10 HPO terms and their corresponding synonyms were sent to three experts who were tasked with pre-selecting the translators by rating the translations on the Likert scale provided. For the ranking, the average Likert Ratings (LR) for each translator were determined. Subsequently, a total of six experts made another assessment for the remaining 120 HPO terms in the second evaluation step. Results were visualized using average LR and standard deviations (SD).

3. Results

The result of the first evaluation round was conclusive. The evaluation of the translators' ratings for the first 10 HPO terms resulted in a unanimous top five ranking: DeepL (LR = 1.22), Bing (LR = 1.51), TC (LR = 1.53), Baidu (LR = 1.70) and Google (LR = 1.73). To compare with the range of all ratings: The next best rated translator is Reverso (LR = 1.95) and the bottom rated translator is Sogou (LR = 3.86).

The results of the second evaluation round are shown in Fig. 3. Translations from DeepL (LR = 1.25) are rated most favorably across all expert ratings, followed by Bing and TC (LR = 1.36). At DeepL, the SD is relatively low for both the 100 random terms (SD = 0.5) and the 20 most common terms (SD = 0.54), which means that there are not many outliers, in other words few individual terms or synonyms received a particularly poor rating. The translations from Baidu (LR = 1.51) and Google (LR = 1.62) are less approved and feature a higher SD, especially for the 100 randomly selected HPO terms.

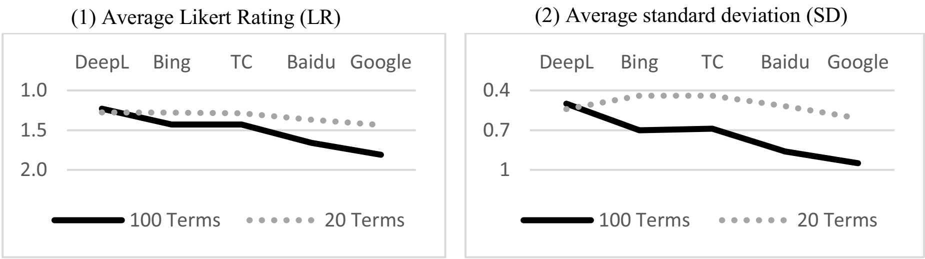

Figure 3. (1) Average Likert Rating of HPO translations (terms + synonyms); (2) Average standard deviation of HPO translations (terms + synonyms). A low value is aspired to for both the mean and the standard deviation.

From the additional comments of the experts, the following sources of translation errors emerged across all translators: spelling errors (i.e., medical terminology), imprecise wording (unspecific or misleading translations), atypical wording (non-medical jargon), incomplete translation (term is not translated or only partially translated) and duplicates (i.e., the translation of different English synonyms results in identical German synonyms).

4. Discussion

This paper investigated the automatic machine translation of ontological concepts from English into German using commercial translation software. In the final evaluation of the HPO translation candidates, the medical experts rated DeepL as the most approved translator with a Likert Rating of 1.25 and a standard deviation of less than 0.55.

A limitation of the methodology described is the strong dependence of the assessments on the individual preferences of the experts. To make a more general statement about the quality of the translator software, the scope of the study would need to be broadened in terms of translation engines, the set of ontological concepts to be translated and the number of medical or translation experts involved in the evaluation.

In this study, translations were done directly from English into German. Translation via multiple translation paths combining different input and support languages would be a way to create more translation candidates for the target language and therefore further optimise machine translation [4]. The additional combination of different translation engines can also be a way of eliminating sources of error such as spelling mistakes, imprecise wording, or incomplete translations.

Another beneficial prospective would be a functional extension of DeepL to include alternative translation suggestions. This function already exists in the DeepL application but has not yet been made available for their API [9]. The expanded availability of translation candidates could avoid duplicate translations in the target language and provide more comprehensive coverage of phenotype synonyms.

In Europe, there are currently a few working groups collaborating on a German representation of English-language ontologies. The University Hospital *Charité* in Berlin is working on a German version of the HPO [2], and the Medical University of Graz is working on such a version for SNOMED CT [4, 10]. However, these and most other projects aiming to make ontologies multilingual are still in the early stages and not very comprehensive but could be impacted and supported by the results of this study.

5. Conclusions

This proof-of-concept has shown that commercial translation software can be useful for translating ontological concepts. However, further optimization and evaluation is needed. The translation of ontologies into multiple languages will be crucial for standardisation and dissemination of medical knowledge and automatic processing of patient data.

Acknowledgements

The author was supported by medical experts from the University Hospital Frankfurt. This study is part of the SATURN project and is funded by the Federal Ministry of Health in Germany (Reference: 2520DAT02B). Authors declare to have no competing interests.

References

[1] Keikhosrokiani, P. Introduction to mobile Medical Information System (mMIS) development. In: Keikhosrokiani, P. editor. Perspectives in the development of mobile medical information systems, San Diego: Elsevier Science & Technology; 2020. p. 1–22, doi: 10.1016/B978-0-12-817657-3.00001-8.
[2] Wołk K, Wołk A. Machine enhanced translation of the human phenotype ontology project. Procedia Comput Sci. 2017 Jan;121:11-8, doi: 10.1016/j.procs.2017.11.003.
[3] Cornet R, Hill C, de Keizer N. Comparison of three english-to-dutch machine translations of SNOMED CT procedures. Stud Health Technol and Inform. 2017 Jan;245:848–52, doi:10.3233/978-1-61499-830-3-848.
[4] Prunott A, Schulz S, Boeker M. Automatic generation of german translation candidates for SNOMED CT textual descriptions. Stud Health Technol and Inform. 2021 May;281:178-182, doi: 10.3233/SHTI210144.
[5] Scott ER, Wallsten RL. A look to the future. In: Lam YWF, Stuart SR, editors. Pharmacogenomics. challenges and opportunities in therapeutic implementation. San Diego: Elsevier Science & Technology; 2019. p. 271–288, doi: 10.1016/B978-0-12-812626-4.00010-3.
[6] Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, Baynam G, Bello SM, Boerkoel CF, Boycott KM, Brudno M. The human phenotype ontology in 2017. Nucleic Acids Res. 2017Jan;45(D1):D865-76, doi: 10.1093/nar/gkw1039.
[7] Human Phenotype Ontology. https://hpo.jax.org/app/download/ontology (accessed on 2022/08/11).
[8] Tse, U. Translators. https://pypi.org/project/translators/ (accessed on 2022/08/11).
[9] DeepL API. https://www.deepl.com/docs-api (accessed on 2022/08/11).
[10] Federal Institute for Drugs and Medical Devices: Terminologies - German translations of SNOMED CT concepts. https://www.bfarm.de/EN/Code-systems/Terminologies/SNOMED-CT/German-Translations/_artikel.html (accessed on 2022/11/18).