

Advantages and pitfalls of machine translation for party research: the translation of party manifestos of European parties using DeepL

Johanna Ida Plenter

Article - Version of Record

Suggested Citation:

Plenter, J. I. (2023). Advantages and pitfalls of machine translation for party research: the translation of party manifestos of European parties using DeepL. *Frontiers in Political Science* , 5, Article 1268320. <https://doi.org/10.3389/fpos.2023.1268320>

Wissen, wo das Wissen ist.



UNIVERSITÄTS- UND
LANDESBIBLIOTHEK
DÜSSELDORF

This version is available at:

URN: <https://nbn-resolving.org/urn:nbn:de:hbz:061-20241030-104736-3>

Terms of Use:

This work is licensed under the Creative Commons Attribution 4.0 International License.

For more information see: <https://creativecommons.org/licenses/by/4.0>



OPEN ACCESS

EDITED BY

Felix Ettensperger,
University of Freiburg, Germany

REVIEWED BY

Jeremy Dodeigne,
University of Namur, Belgium
Michele Scotto Di Vettimo,
University of Exeter, United Kingdom

*CORRESPONDENCE

Johanna Ida Plenter
✉ johanna.plenter@hhu.de

RECEIVED 27 July 2023

ACCEPTED 31 October 2023

PUBLISHED 22 November 2023

CITATION

Plenter JI (2023) Advantages and pitfalls of machine translation for party research: the translation of party manifestos of European parties using *DeepL*.
Front. Polit. Sci. 5:1268320.
doi: 10.3389/fpos.2023.1268320

COPYRIGHT

© 2023 Plenter. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages and pitfalls of machine translation for party research: the translation of party manifestos of European parties using *DeepL*

Johanna Ida Plenter*

Department of Social Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Parties are the central actors in representative democracies as they perform important democratic functions. Thus, the identification of party positions is a crucial concern. Party researchers mainly rely on party manifestos to estimate policy positions. However, the analysis of manifestos is accompanied by challenges—one of the biggest being cross-national comparisons because of different institutional settings and languages. This article discusses machine translation (MT) as a new option for party research, and reports on the author's experiences with the translation of more than 200 party manifestos using the commercial artificial intelligence (AI) translation tool *DeepL*. To make this approach widely applicable, the (technical) procedure, including its problems and workarounds for large-scale projects, is presented as a step-by-step guide using R. Additionally, drawing on the most recent German, Estonian, Italian and Polish parliamentary election manifestos this article evaluates the quality of the *DeepL* translations by applying both back translation and Wordfish analyses. The main findings indicate that *DeepL* offers high-quality translations as more than 90% of the checked sentences are reproduced word-for-word or at least synonymously and with stable positioning on the left-right scale of both original and English translation. The results have greater implications for political science research as they speak to the reliability of machine translation for political texts.

KEYWORDS

party manifestos, machine translation, text as data, *DeepL*, translation quality

1 Introduction

Parties are the central actors in representative democracies and electoral competition as they perform important democratic functions such as aggregating and articulating citizens' interests. To understand electoral processes or estimate party responsiveness, the identification of party positions is a crucial concern in electoral and party research—in fact, it has become a subdiscipline in its own right (Laver, 2001). There is a wide range of methodological approaches and types of data to determine party positions: they can be estimated with the help of expert or mass surveys, legislative voting behavior, media analyses, or based on texts (for a detailed discussion, see Mair, 2001). Within the text-based methods, a further distinction can be made between quantitative and qualitative analysis methods. In addition, text analyses can draw on various types of text data, e.g., parliamentary or candidate speeches (Lauderdale and Herzog, 2016; Atzpodi, 2020) as well as policy papers or coalition agreements (Benoit et al., 2005; Gross and Krauss, 2021). More frequently,

however, election or party manifestos¹ are relied upon to estimate parties' policy positions (Slapin and Proksch, 2008; Volkens et al., 2013; Bräuninger et al., 2020).

Election manifestos are so well suited for an analysis of party positions as they are a fairly reliable and readily available data source. However, their analysis is also accompanied by challenges—one of the biggest being cross-national comparison because of the multitude of languages (Lucas et al., 2015, p. 255). Especially if the manifestos will be coded manually, either native-language coders/analysts have to be hired or the texts have to be translated.² When it comes to translation, there are two options: professional translators or machine translation (MT). The latter method and its pros and cons will be examined in this article to identify opportunities and obstacles for the estimation of party and policy positions based on automatically translated manifestos. The article primarily concerns the MT tool *DeepL* and based on manifestos in four European languages it evaluates its translation quality using back translation and Wordfish analyses. The main results of this evaluation seem to confirm the company's promise of excellent quality: first, the back translation shows that more than 90% of the checked sentences are reproduced word-for-word or at least synonymously. Second, the Wordfish analysis shows that the positioning of the automatic translation on the left-right scale is very close to that of the original text indicating that the translation did not change the tone, content, and political implications.

The remaining article is structured as follows: first, I discuss current developments and state-of-the-art MT methods to outline their benefits and shortcomings compared to human translation. Afterwards, I exemplify this by presenting a step-by-step guide to implementing MT with the commercial artificial intelligence (AI) tool *DeepL*³ for large-scale projects. The added value of this section is that an evaluation of the translations is carried out to assess their quality as an attempt to open the black box of the AI machine translation algorithm. The subsequent conclusion summarizes the results and evaluates them against the backdrop of existing limitations.

2 Machine translation: state-of-the-art methods in comparison to human translation

In translation studies, machine translation is defined as “the automatic conversion of text from one natural language to another”

and it is the “process in which the interlingual conversion of text is carried out by a machine, even if the proper functioning of that machine relies on human labor before, during or after run time” (Kenny, 2019b, p. 305–306). Even though machine translation has been used for several decades, the technical developments of the past years have led to a great surge of innovation, for example, the development of end-to-end neural machine translation (NMT), which has become the state-of-the-art method (Tan et al., 2020, p. 5). Neural machine translation is what is meant when we generally talk about translation with the help of AI or deep learning. It is neither the aim nor within the scope of this article to trace the development stages and technical aspects of MT,⁴ but rather to discuss its advantages for political science and especially party research. Nevertheless, it is worth briefly outlining its simplified functionality: essentially, NMT works in the interaction of source to target language, in that the neural network predicts the translation word by word using word embeddings, i.e., the spatial representation of words that exhibit the semantic relationships among them. With increasing amounts of training data, the network “learns” and, thus, improves the quality of the translations. However, for commercial MT algorithms it is difficult to assess robustness and interpretability because these models are essentially black boxes as we know neither their training data nor coding. From Tan et al. (2020, p. 16), it can be concluded that “noisy inputs” i.e., erroneous spelling or incorrect usage of words, are a particular problem to the translation's robustness. Yet, it can be assumed that manifestos and other official documents published by political parties are likely to be almost free of errors, as they are usually well-curated and edited. This is one reason why, for translated text data, manifestos are superior to e.g., parliamentary speeches.

Machine translation has some decisive advantages and disadvantages compared to “traditional” translation by humans. First, one of the most severe disadvantages is undoubtedly the fact that AI has no feeling for language and does not register sensitive language usage—especially if context is lacking. Examples of this are irony and sarcasm as well as idioms or (political) expressions established at a certain point in time, such as the NATO's 2% defense investment guideline, which in German is commonly referred to as the “Zwei-Prozent-Ziel” (literal translation: “two-percent-target”). As it is currently discussed, the term is well-known among policy-makers, the media, and the informed public, but 20 years from now that might not be the case—making the term meaningless. Related to this point is a second drawback of MT, namely its handling of language change or evolution. That languages change over time is a normal and inherently unproblematic process. For MT, however, semantic change in particular can be a difficulty because the usage and connotation of certain terms change. In English, for example, the use of the word “gay” has changed at least three times (Lalor and Rendle-Short, 2007, p. 148; Shi and Lei, 2020, p. 35). Accordingly, the expression was first used with a positive connotation as a synonym for “jolly” or “happy”, this changed to the neutral

1 In this article, the expressions “election manifesto”, “election program”, and “party manifesto” are used interchangeably and synonymously. Existing differences between these terms (see Klingemann et al., 1994, Chapter 2 for a discussion) are not relevant for this paper.

2 If the data is to be processed quantitatively using machine learning and the results do not need to be validated or coded manually, recent developments in computational linguistics show that multilingual classifiers e.g. using multilingual sentence embeddings provide valid results (see e.g., Licht, 2023). Accordingly, no translation is necessary for these types of applications.

3 *DeepL* is available at www.deepl.com; corporate headquarters of *DeepL* SE are in Cologne, Germany.

4 See Kenny (2019a) for a detailed discussion of the historical developments of machine translation and/or Tan et al. (2020) for a review of NMT developments, methods, and tools.

meaning “homosexual”, and more recently to a negatively connoted expression for “boring” or “lame”. A similar development can be observed with regard to the “N-word” and the linguistic representation of Black people in general (Washington, 2023). Even though MT’s handling of such cases has not yet been systematically evaluated in the literature, it can be assumed that incorrect or offending translations may occur—especially if the corpus spans several decades, i.e., potentially across semantic changes. As human translation still marks the gold standard, it can be expected that professionally trained translators better capture such language use because of their context knowledge. Bizzoni and Lapshinova-Koltunski (2021), however, find that translations from different translators are stylistically quite heterogeneous. This implies that the translation quality is highly dependent on the individual and their language and content knowledge. However, further research that investigates the comparability of different human translators is needed. Alternatively, computational linguistics research discusses the possibility of combining MT with human translation in order to use “the best of both worlds” (Li et al., 2023, p. 9511; Peña Aguilar, 2023). In this approach, human translations are fed into the training dataset of the MT algorithm to improve its quality. Accordingly, this approach requires access to the MT algorithm, which—at least for commercial software—is usually not available. Lastly, there may also be quite mundane obstacles when using machine translation, e.g., the preferred MT does not offer all languages or is not available in a country. The opacity of AI tools described above also prevents outsiders from assessing how well which tool has been trained with which inputs and, above all, languages and language combinations. Since, however, these obstacles cannot be eliminated by the researcher(s), they must at least be considered as a limitation in a critical reflection.

On the other hand, one extremely significant advantage of MT compared to human translation is its resource efficiency. This is to be understood in two respects: for one, machine translation requires significantly less time, and for another, financial resources. This fact makes MT particularly interesting for research projects that have to get by without generous funding and/or a large team and must be completed within a foreseeable schedule. Another advantage of machine translation lies in the deep-learning approach of neural translation tools as with every translation, i.e., data input, they are continuously learning and improving. Consequently, MT algorithms draw on a constantly growing body of knowledge, whereas typically only one translator is hired per document, meaning that translation quality depends on that person.

In summary, when deciding between machine and human translation, the potentially poorer or more error-prone translation quality must be weighed against the significant time and money savings of real-time translation. For large corpora, human translation is just not feasible because of the sheer amount of text. In addition, the type of text to be translated and the time frame in which it was created are also decisive. For example, irony and sarcasm, which can be hard to grasp for a MT algorithm, play a much greater role in political speeches than in press releases and party manifestos. If the time period in which the texts were created spans several decades, it should also be reflected to what extent language change could influence the results. Moreover, Reber (2019, p. 118) points out that the “choice of method”, i.e.,

full-text translation vs. translation of individual words/expressions, must be considered because MT algorithms rely on the context in which a word is used. He concludes that the translation of entire documents is more accurate and should, thus, be the first choice. Accordingly, a blanket recommendation for one or the other does not make sense; however, especially given further development, training, and evaluation of MT, it is expected that this method will become more and more established in (political science) research.

3 Machine translation with *DeepL*

Unsurprisingly, machine translation has already been applied in political and communication science research and has also been evaluated (i.e., Lucas et al., 2015; de Vries et al., 2018; Düpont and Rachuj, 2022). Furthermore, Lucas et al. (2018) provided the R package *translateR*, which enables translation with API integration. All of these applications have used Google Translate for the translation.⁵ This article, in contrast, hereafter presents the application of the commercial AI *DeepL* and discusses the advantages of this MT tool. For this purpose, I will report on my experiences with translating current European manifestos. In addition to manifestos, there are many other possible applications of MT in party research, such as party statutes, social media posts, or speeches. However, the above-mentioned considerations and potential shortcomings must be assessed for different types of textual data. Before presenting the procedure, first, some brief explanations about *DeepL*.

DeepL is a translation AI by the same-named German company, which has been available since 2017 for an initial seven (exclusively European) languages. At the time of writing (July 2023), *DeepL* offers translations for 31 languages, which, according to the company’s statements, significantly exceed the quality of other machine translations. This quality is said to have been evaluated both by “scientific tests” and “external professional translators” (DeepL, 2022). However, these claims are difficult to verify because the sources of the tests and evaluations are neither published, provided, nor cited on the company’s homepage. Yet, several technology magazines and media outlets generally confirm these statements (i.e., Coldewey and Lardinois, 2017; Wyndham, 2021). Additionally, first academic assessments seem to confirm the good translation quality. Hidalgo-Ternero (2020, p. 170) compares Google Translate and *DeepL* translations of Spanish idioms and concludes: “[O]verall, DeepL slightly outperforms Google Translate [...] [as] the global results exhibit an accuracy rate of 70% [...] for Google Translate and 78% [...] for DeepL.” Also focusing on the Spanish-English language combination, Peña Aguilar (2023) finds that *DeepL* outperforms both Bing and Google Translate. Lastly, in his comparison of Google Translate and *DeepL*, Reber (2019, p. 117) concludes that both tools perform equally well for full-text translation. For my research, I chose *DeepL* partly because of this said translation quality. More importantly however, it was decisive that *DeepL*—in contrast e.g., to Google Translate or Amazon Translate—can process plain text files (txt format), which

⁵ With *translateR* you can choose between the Google Translate API and the Microsoft Translator API.

are the standard file format especially in scientific quantitative text analysis.

DeepL offers multiple usage options tailored to different needs. For one, there are both free and paid subscriptions; for another, *DeepL* features simple text translation in the web or app translator, translation of entire files and API integration. The *DeepL* API can also be integrated into R using the *deeplr*-package (Zumbach and Bauer, 2021), which is a good alternative to text or file translation, in particular when the text to be translated is not exceedingly long. Up to 500,000 characters per month (~280 standard pages) can be translated free of charge; in addition to a monthly fee, API Pro is billed according to actual consumption. So before deciding for or against API translation, it should be calculated how many characters the text to be translated comprises. In the following, I will concentrate exclusively on file translation which is offered for six file formats (docx, htm, html, pdf, pptx, txt), but as this guideline aims to make the translated manifestos accessible for (automated) text analysis, which requires machine-readable data, only text files (txt) are considered here. The code provided in the [Supplementary material](#), however, also includes API translation (see Reber, 2019 for an exemplary application of the *DeepL* API).

Summarizing, again no general and universally valid recommendation can be made with regard to the selection of the MT tool. Rather, the decisive factors are the volume of the text data to be translated, the file format in which it is available, and the analyses that are to be performed following translation. For the aforementioned reasons, however, this article focuses only on *DeepL*.

3.1 Step-by-step implementation with R

Turning to the data preparation for translation: after collecting all manifestos (or other documents) to be translated, in a second step, the files have to be converted to txt. One of the biggest challenges of file conversion, however, is that manifestos—more than party statutes or speech manuscripts—often exhibit an elaborate layout or include figures, pictures, quotes, and tables. Text recognition and extraction are particularly challenging when the text is set in columns. When converting such a file from pdf to txt, a lot of text can get lost or mixed up as it is read in lines and not columns. It would therefore be highly desirable if parties made all of their communication available as plain text documents and if database projects such as the Manifesto Project or OPTED⁶ were further strengthened and funded. The current version of the Manifesto Corpus (Lehmann et al., 2023), available through the R package *manifestoR* (Lewandowski et al., 2020), already contains more than 3,000 machine-readable manifestos, so no separate file conversion is necessary for these. For file conversion of the additional election programs, I used the *tabulizer*-package (Leeper, 2018) because it is capable to recognize text set in columns (see RScript in the [Supplementary material](#) for instructions and code).

As a third step, all converted files should be checked to see whether the text recognition has worked without errors or whether

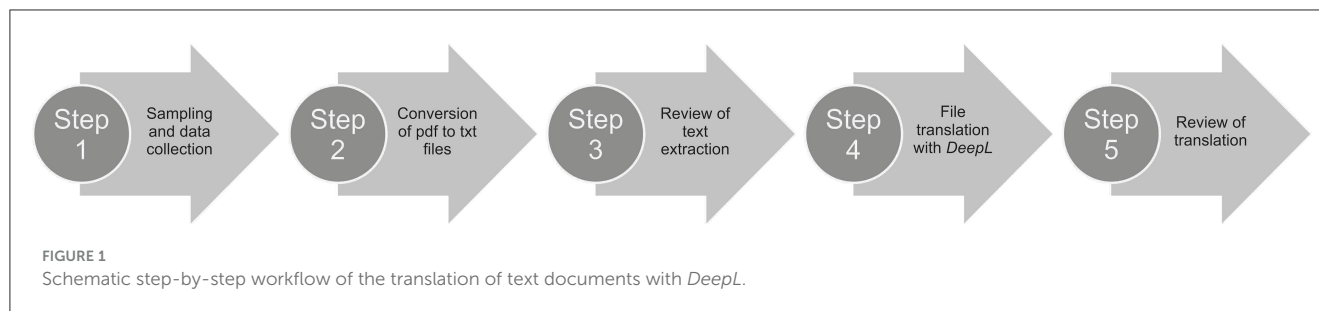
corrections need to be made. It appears that non-embedded fonts and special characters (i.e., not UTF-8 encoded) are problematic because the AI may not identify them as text (see Lucas et al., 2015, p. 256–257 for a detailed discussion). Depending on the amount of unrecognized text, manual post-processing of the files is possible but time-consuming. For the translation, the converted files must now be exported from R as txt files. The file translation function of *DeepL* can process several files simultaneously in both the browser and the app so the actual translation happens in real time. After translation, it is advisable to again check whether the entire file/text has been correctly recognized and translated. [Figure 1](#) below briefly illustrates the workflow step by step.

Summarizing, it can be said that MT in general and *DeepL* in particular offer great benefits for party research, since the fast and cheap machine translation of e.g., manifestos makes it easy to analyze the entire text and its message. This enables statements about policy positions and issues that are not included in the Manifesto Project codebook or about the sentiments, i.e., framing, parties use. The biggest technical challenges in dealing with AI are certainly file format as well as the preparation and post-processing of the documents. Overall, the advantages outweigh the shortcomings and considering the self-learning infrastructure, it can be assumed that machine translation algorithms will further improve in the future.

3.2 Evaluation of *DeepL* translation quality

One of the biggest issues in translation studies is the assessment and evaluation of translation quality. Since the entire value of a translation hinges upon its quality, it is important to develop certain measurements and discuss their respective strengths and weaknesses. It should however be noted that the “concept of quality varies greatly, within and outside translation studies” (Colina, 2008, p. 98), which is why it is essential to define the aim of the evaluation beforehand. In the context of political science research, it is crucial to determine to what extent the original and the translation correspond in terms of content. Word usage and tonality are central, especially when texts are used to determine party positions or framing. It can be argued that most political communication is drafted and tailored to the author’s agenda irrespective of whether a party or a single candidate communicates. Thus, it is essential that the translation quality, i.e., the degree to which original and translation correspond, is extremely high. For this reason, in the following, I evaluate the quality of *DeepL* translations with two purely descriptive assessment approaches, namely back translation and Wordfish analyses. To this end, I draw on a subset of the Manifesto Corpus that comprises the most recent manifestos of parties from Germany, Estonia, Italy, and Poland. These four cases were selected for two reasons: first, the languages mainly spoken in these countries belong to different branches of the Indo-European or Uralic language family. While German as a Germanic language, Italian as a Romance language and Polish as a Slavic language are branches of the Indo-European language family, Estonian as a Finnic language belongs to the Uralic language family. Second, the four languages differ extremely in terms of their prevalence as measured by the number of native

⁶ For more information on the Manifesto Project see: <https://manifesto-project.wzb.eu/> and for OPTED see: <https://opted.eu>.



speakers. According to the Ethnologue, German is spoken as a native language by about 75 million people, Italian by 65 million, Polish by 40 million and Estonian by 1.2 million (Eberhard et al., 2023). Thus, the evaluation of the translation quality does not only look at one language family or branch, but takes several combinations into account. Additionally, it concerns both very common (i.e., German—English, Italian—English) and rarer (i.e., Polish—English, Estonian—English) language combinations. It seems plausible to assume that the language prevalence is correlated to the number of data inputs, i.e., the training, of these language combinations (de Vries et al., 2018, p. 419). However, since the algorithm is unknown to outsiders, this is purely speculative for DeepL.

The first evaluation uses back translation, also called re-translation, and only draws on the seven German manifestos. As the name indicates, back translation takes translated texts and translates them back to their source language to compare them with the original. Behr (2017) discusses the pros and cons of this method and concludes that back translation is very straightforward and, especially with machine translation, fast. However, the method lacks a clear conception of what is considered a mistake or “poor quality”. In her empirical analysis, Behr (2017, pp. 581–582) finds that back translation “can successfully identify errors [...]”; however, most of these issues were identified by actual translation assessment as well”. She, therefore, concludes that this method should always be accompanied by other assessment approaches. Despite its shortcomings, back translation is still one of the most widely used methods for evaluating translation quality. For this reason, it is applied in this article as a first step.

To estimate the quality of the DeepL translations, I draw on the seven manifestos by German parties and re-translate them from English to German—again using DeepL. For each of these manifestos, 80 sentences of the back translation are manually coded and compared to the original. The manual coding classifies each sentence as either verbatim back translation or synonymous back translation; in addition, obvious errors are also marked.

Looking at Table 1, the most central result shows that on average roughly 21% of the sentences were re-translated in verbatim form. Additionally, it seems as if the back translation produces reliable results with an average of only 3.4 mistakes per document. The translations were coded as faulty when there was an evident change in meaning or when a technical error, e.g., double translation of a word/phrase, occurred. Most of the sentences, however, have been re-translated as a paraphrase of the original which reflects the content synonymously. Remarkably, both the number of errors and the share of verbatim sentences vary quite

TABLE 1 Results of the back translation of German manifestos.

Party	No. of errors	Verbatim sentences (in %)
AfD	4	17.50
CDU/CSU	3	25.00
FDP	2	26.25
SPD	4	20.00
Greens	3	20.00
SSW	3	18.75
The Left	5	22.50
Average	3.4	21.43

The bold values are column averages.

significantly between the seven manifestos. One reason for this might be that the parties use sentences of different lengths and, thus, complexity. The probability of re-translating short sentences verbatim is higher than for long sentences. Nevertheless, it can be stated that based on this exemplary check of the manifestos of German parties, the DeepL translations can be assessed as being of high quality. On average, more than 90% of the sentences were re-translated either word-for-word or synonymously, indicating a good quality of the initial German-to-English-translation.

In order to check whether the back-translation changes the text content, a dictionary analysis was also performed. For this purpose, the seven German manifestos were compared both in the original and the back-translated version by searching the texts for four concepts: economy, state, environment, and social affairs. The results are presented in Table 2. Overall, the results confirm the good translation quality, as the number of hits in the original and the back translation are very similar. The largest deviations are found within the concept of social issues in the manifestos of the Greens (+9 hits) and the Left (+19 hits). In six cases, however, exactly the same number of hits was found in both documents. Therefore, the results of this small dictionary analysis can be interpreted as indicating that manifestos translated with the help of DeepL are suitable for empirical text analyses because they do not substantially change the text content.

In contrast to the back-translation approach, the second assessment method looks at the manifestos as a whole. To accompany the back translations, a Wordfish estimation was conducted. Wordfish is an unsupervised scaling technique developed by Slapin and Proksch (2008) that positions texts onto a one-dimensional scale, e.g., the left-right scale. The model does not

TABLE 2 Comparison of dictionary analysis of original and re-translated German manifestos.

Party	Economy		State		Environment		Social affairs	
	Original	Back translation	Original	Back translation	Original	Back translation	Original	Back translation
AfD	17	20	24	27	8	6	33	30
CDU/CSU	46	43	54	54	8	9	42	47
FDP	25	27	37	37	6	6	21	20
SPD	16	18	10	10	10	9	47	48
Greens	42	42	45	49	28	33	78	87
SSW	12	10	9	9	9	8	26	25
The Left	51	57	21	20	25	28	206	225

require reference texts or other previous information but instead uses word frequencies, thus the parties' relative word usage, to place texts along this dimension and it assumes a Poisson distribution. To my knowledge, Wordfish has not yet been applied as a translation assessment method—probably because it was not designed for this purpose. However, I argue that it can be useful for political texts in particular because all (party-) political communication is carefully drafted to convey only the intended message. Thus, I argue that the translation quality can be determined based on the correspondence or distance between the positions of the original and the translation. To assess the quality of the translation, I estimate and compare the relative positions of original (German/Estonian/Italian/Polish) and the translated manifestos (English). [Figure 2](#) shows the results of the Wordfish placement of all manifestos.

At first glance, it is already apparent that the positioning of the originals and their translations vary in proximity depending on the language combination. While the combinations with Estonian, German and Italian largely produce highly congruent positionings, the Polish-English combination in particular seems to perform more poorly. Thus, the proximity of positions does not seem to depend on language family or language prevalence, i.e., number of native speakers. Overall, the result for all four combinations can be considered (very) satisfactory. The second important result of the review is that positional deviations can be observed for all languages and party families, both to the left and to the right. In each tested combination, about 50% of the translations deviate slightly to the left and 50% deviate slightly to the right—regardless of whether the party itself is left or right. This indicates that there is no systematic left or right bias in the translations and/or language combinations.

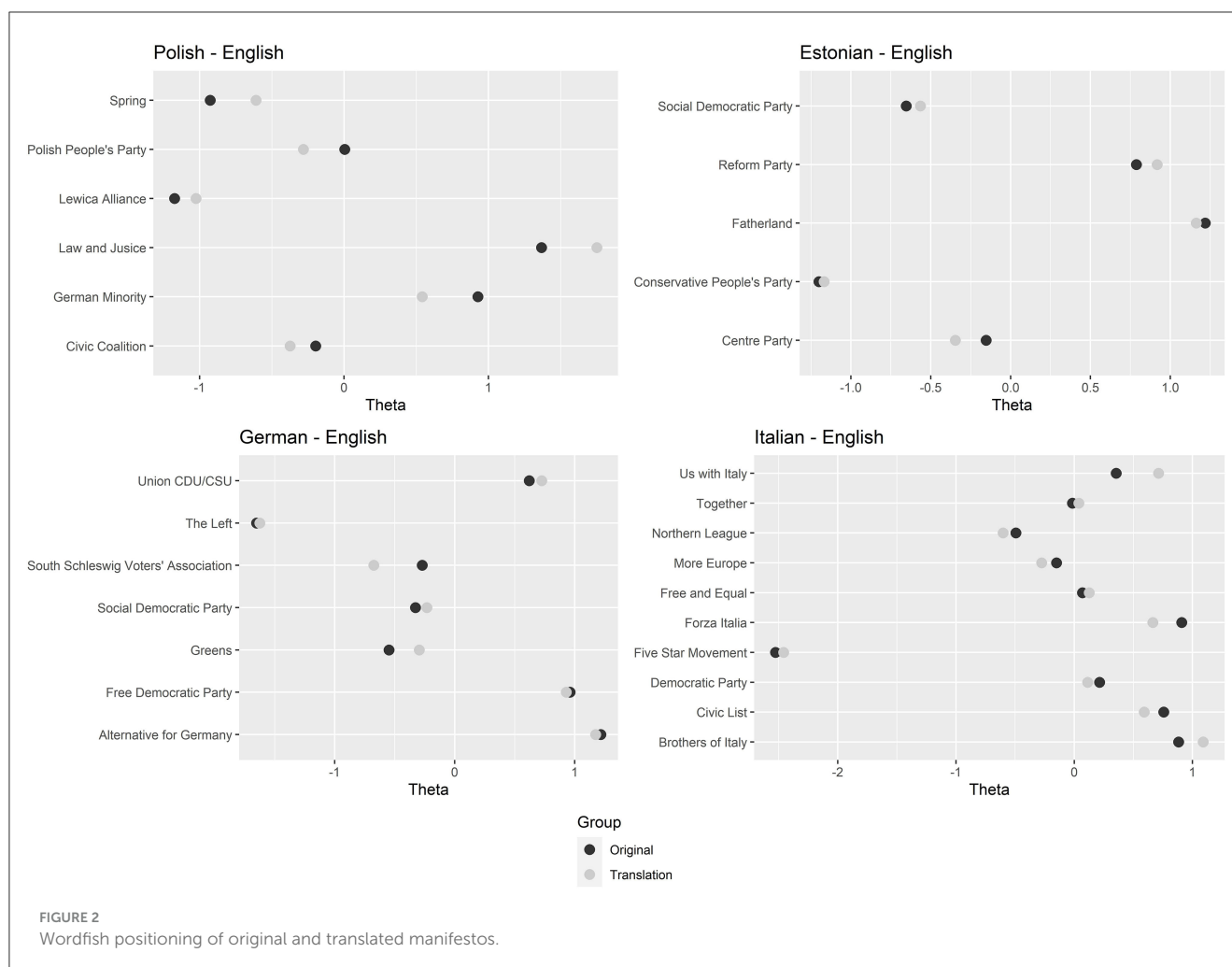
One explanation for the deviations could be the significantly varying number of so-called document features in the respective languages. Due to linguistic differences between German, Polish, Estonian and Italian on the one hand and English on the other, this number is extremely reduced between original and translation. Consequently, the Wordfish algorithm has fewer word frequencies to rely on, implying that every feature becomes more influential. Overall, the Wordfish analysis confirms the results of the back-translation approach. As the placement of the original and translated manifesto is very similar, it can be concluded that the *DeepL* translation did not change the content of the manifestos and their political implications. Nevertheless, it must be mentioned at this point that neither back translation nor Wordfish allow a

statement about the correspondence of the framing or the mood in the original and the translation.

The extent to which the results discussed here are transferable to other types of text depends on several factors and cannot be universally answered. On the one hand, the characteristics of the (political) text type play a role. In contrast to speeches, manifestos are considered “sober” ([Hawkins and Castanho Silva, 2018](#), p. 31), since they address a different audience and must represent the interests of the entire party. It is therefore plausible to assume that this characteristic, reflected e.g., in the use of irony in the text, influences translation quality. On the other hand, the length of the texts and within them the length and linguistic complexity of the sentences and words are also decisive. These factors typically vary between types of political texts ([Tolochko and Boomgaarden, 2019](#)). For example, due to the character limitations alone, tweets are significantly less complex and exhibit different language usage than formal party communications—a fact that can also affect translation quality. Thus, all of these features and distinctions must be considered when transferring machine translation to other textual data.

4 Conclusion

The overall aim of this article was to discuss the benefits and pitfalls of machine translation, particularly using *DeepL*, for political science and party research. I present the process of automatically translating a large set of party manifestos with its difficulties and workarounds using R and *DeepL*. For large-scale cross-national studies, multilingual text data are a challenge that can be met using MT. The step-by-step guide presented here will assist other scholars with such projects, and the commercial translation AI *DeepL* offers a solution by providing high-quality translations for 31 languages. The most tedious challenge within the workflow is the pre-processing of files, i.e., the text extraction, as manifestos exhibit an increasingly sophisticated layout and have grown in length. In general, it can be stated that the more complex the layout, i.e., the typesetting, of the source file, the more difficult the text extraction. To provide further added value to the technical implementation, in a second step I tried to open the black box of the MT algorithm a bit to evaluate the quality of the machine-translated texts. After all, translation quality and



thus the tool's scientific trustworthiness is of central importance, especially for the analysis of party positions and rhetoric. Therefore, I performed an exemplary evaluation of the *DeepL* translations both by re-translating a set of manifestos of German parties and by determining the relative position on the left-right scale of German, Estonian, Polish and Italian manifestos using Wordfish analyses. The results of this evaluation indicate that *DeepL* offers high-quality translations, which do not significantly change the content and the positioning. The back translation revealed that *DeepL* re-translates 90% of the sentences either word-for-word or at least synonymously with an average error rate of 3.4 errors per 80 sentences. The subsequent Wordfish analysis showed for all language combinations that the relative positions of original and translation are very close and that there is no systematic right-left bias. Nevertheless, it should not be forgotten that most MT algorithms are black boxes that cannot be fully opened by the evaluation carried out here. Both future research and further development of the algorithms would be necessary to make MT an integral part of social science research.

In contrast to the approach presented here, there are recent developments to process and analyze multilingual corpora in their original language. New transformer models such as BERT (Devlin et al., 2019) or multilingual sentence embeddings (Licht, 2023)

were enabled by advances in computational linguistics in the field of large language models (LLMs) and provide valid and reliable results. Which of the two approaches should be chosen in each individual case depends centrally on how the text data are to be processed in the further course of the research project. The central advantage of translated text data is that, on the one hand, the entire corpus can be examined in one analysis, e.g., a machine learning algorithm, and, on the other hand, that the results can be inspected by one researcher and further examined, e.g., with manual coding. This paper contributes to the validation of analyses using translated text data, as the main findings show that MT—and particularly *DeepL*—produces reliable and trustworthy results. The importance of this contribution is also underlined by the fact that, according to de Vries et al. (2018, p. 418), many authors either simply assume that MT provides reliable results or do not consider this issue at all.

While this article concentrates on party manifestos and policy positions, MT can be valuable for other types of textual data and analyses as well. In fact, the amount and availability of (political) textual data such as press releases, speeches, newspaper articles, and social media posts is growing every day and MT is one way to make this data accessible for automated comparative research. The added value of machine translation in general and *DeepL* in particular is that the entire text, including its framing and rhetoric, is made

accessible for analysis. This allows for broader research perspectives that cannot be covered, for example, by existing coding schemes such as the Manifesto Project.

Concluding, all of this leaves us with the scientific-ethical question of whether such a further “algorithmization” of (social science) research is desirable and a trend worth supporting. Particularly in light of open science initiatives, this development must be critically questioned. This applies all the more to commercial software and algorithms like *DeepL*. In his opinion commentary, Arthur Spirling (2023, p. 413) therefore claims: “The rush to involve such artificial-intelligence (AI) models in research is a problem. Their use threatens hard-won progress on research ethics and the reproducibility of results. Instead, researchers need to collaborate to develop open-source LLMs that are transparent and not dependent on a corporation’s favors.” Consequently, it would be extremely desirable for companies to disclose their algorithms or for open-source MT tools to be developed that provide high-quality translations for a variety of language combinations. For such tools, it would then also be possible to combine human and machine translation—an approach that promises significant quality improvements of up to 28% (Li et al., 2023, p. 9512). Until such tools are available, however, I argue in this article that commercial AI tools such as *DeepL* are a good alternative. After all, MT tools can also contribute to making science more accessible and inclusive by enabling the analysis of countries that are not typically the focus of interest. Furthermore, machine translation enables scientists without large financial resources to participate in and contribute to the research on political parties.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

References

- Atzpodi, D. S. (2020). Party competition in migration debates: the influence of the AfD on party positions in German state parliaments. *Ger. Polit.* 31, 381–398. doi: 10.1080/09644008.2020.1860211
- Behr, D. (2017). Assessing the use of back translation: the shortcomings of back translation as a quality testing method. *Int. J. Soc. Res. Methodol.* 20, 573–584. doi: 10.1080/13645579.2016.1252188
- Benoit, K., Laver, M., Arnold, C., Pennings, P., and Hosli, M. O. (2005). Measuring national delegate positions at the convention on the future of Europe using computerized word scoring. *Eur. Union Polit.* 6, 291–313. doi: 10.1177/1465116505054834
- Bizzoni, Y., and Lapshinova-Koltunski, E. (2021). “Measuring translationese across levels of expertise: are professionals more surprising than students?” in *Proceedings of the 23rd Nordic Conference on Computational Linguistics*, eds S. Dobnik, and L. Øvrelid (Linköping: Linköping University Electronic Press), 53–63.
- Bräuninger, T., Debus, M., Müller, J., and Stecker, C. (2020). *Parteienwettbewerb in den deutschen Bundesländern*. Wiesbaden: Springer VS. doi: 10.1007/978-3-658-29222-5
- Coldewey, D., and Lardinois, F. (2017). *DeepL Schools Other Online Translators with Clever Machine Learning*. San Francisco, CA: TechCrunch.
- Colina, S. (2008). Translation quality evaluation: empirical evidence for a functionalist approach. *Translator* 14, 97–134. doi: 10.1080/13556509.2008.10799251
- de Vries, E., Schoonvelde, M., and Schumacher, G. (2018). No longer lost in translation: evidence that google translate works for comparative bag-of-words text applications. *Polit. Anal.* 26, 417–430. doi: 10.1017/pan.2018.26
- DeepL (2022). *Presseinformationen*. Available online at: <https://www.deepl.com/press.html> (accessed March 11, 2022).
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). “BERT: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, eds J. Burstein, C. Doran, and T. Solorio (Stroudsburg, PA: Association for Computational Linguistics), 4171–4186.
- Düpont, N., and Rachuj, M. (2022). The ties that bind: text similarities and conditional diffusion among parties. *Br. J. Polit. Sci.* 52, 613–630. doi: 10.1017/S0007123420000617
- Eberhard, D. M., Simons, G. F., and Fennig, C. D. (eds) (2023). *Ethnologue: Languages of the World: Twenty-sixth edition*. Dallas, TX: SIL International.
- Gross, M., and Krauss, S. (2021). Topic coverage of coalition agreements in multi-level settings: the case of Germany. *Ger. Polit.* 30, 227–248. doi: 10.1080/09644008.2019.1658077
- Hawkins, K. A., and Castanho Silva, B. (2018). “Textual analysis: big data approaches,” in *The Ideational Approach to Populism: Concept, Theory, and Analysis*, eds K. A. Hawkins, R. E. Carlin, L. Littvay, and C. Rovira Kaltwasser (London: Routledge), 27–48. doi: 10.4324/9781315196923-2

Author contributions

JP: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The publication of this research was funded by the HHU Open Access Fund of the University and State Library Düsseldorf.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpos.2023.1268320/full#supplementary-material>

- Hidalgo-Tertero, C. M. (2020). Google translate vs. DeepL: analysing neural machine translation performance under the challenge of phraseological variation. *MonTI* 154–177. doi: 10.6035/MonTI.2020.ne6.5
- Kenny, D. (2019a). “Machine translation,” in *The Routledge Handbook of Translation and Philosophy*, eds P. Rawling and P. Wilson (London, NY: Routledge), 428–445. doi: 10.4324/9781315678481-27
- Kenny, D. (2019b). “Machine translation,” in *Routledge Encyclopedia of Translation Studies*, eds M. Baker, and G. Saldanha (London: Routledge), 305–310. doi: 10.4324/9781315678627-65
- Klingemann, H.-D., Hofferbert, R. I., and Budge, I. (1994). *Parties, Policies, and Democracy*. Boulder: Westview Press.
- Lalor, T., and Rendle-Short, J. (2007). “That’s So gay”: a contemporary use of gay in Australian English. *Aust. J. Linguist.* 27, 147–173. doi: 10.1080/07268600701522764
- Lauderdale, B. E., and Herzog, A. (2016). Measuring political positions from legislative speech. *Polit. Anal.* 24, 374–394. doi: 10.1093/pan/mpw017
- Laver, M. ed. (2001). *Estimating the Policy Position of Political Actors*. London: Routledge.
- Leeper, T. J. (2018). *tabulizer: Bindings for Tabula PDF Table Extractor Library: R package version 0.2.2*. Available online at: <https://github.com/ropensci/tabulizer>
- Lehmann, P., Franzmann, S., Burst, T., Lewandowski, J., Matthieß, T., Regel, S., et al. (2023). *Manifesto Corpus. Version: 2023-1*. Berlin.
- Lewandowski, J., Merz, N., and Regel, S. (2020). *manifestoR: Access and Process Data and Documents of the Manifesto Project: R package version 1.5.0*. Available online at: <https://github.com/ManifestoProject/manifestoR>
- Li, Z., Qu, L., Cohen, P., Tumuluri, R., and Haffari, G. (2023). “The best of both worlds: combining human and machine translations for multilingual semantic parsing with active learning,” in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), eds A. Rogers, J. Boyd-Graber, and N. Okazaki (Stroudsburg, PA: Association for Computational Linguistics), 9511–9528. doi: 10.18653/v1/2023.acl-long.529
- Licht, H. (2023). Cross-lingual classification of political texts using multilingual sentence embeddings. *Polit. Anal.* 31, 366–379. doi: 10.1017/pan.2022.29
- Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., Tingley, D., et al. (2015). Computer-assisted text analysis for comparative politics. *Polit. Anal.* 23, 254–277. doi: 10.1093/pan/mpu019
- Lucas, C., Tingley, D., and Dehiya, V. (2018). *translateR: R package version 2.0*. Available online at: <https://github.com/ChristopherLucas/translateR>
- Mair, P. (2001). “Searching for the positions of political actors: a review of approaches and a critical evaluation of expert surveys,” in *Estimating the Policy Position of Political Actors*, ed. M. Laver (London: Routledge), 10–30. doi: 10.4324/9780203451656_chapter_2
- Peña Aguilar, A. (2023). Challenging machine translation engines: some Spanish-English linguistic problems put to the test. *Cad. Trad.* 43, 1–26. doi: 10.5007/2175-7968.2023.e85397
- Reber, U. (2019). Overcoming language barriers: assessing the potential of machine translation and topic modeling for the comparative analysis of multilingual text corpora. *Commun. Methods Meas.* 13, 102–125. doi: 10.1080/19312458.2018.1555798
- Shi, Y., and Lei, L. (2020). The evolution of LGBT labelling words: tracking 150 years of the interaction of semantics with social and cultural changes. *Engl. Today* 36, 33–39. doi: 10.1017/S0266078419000270
- Slapin, J. B., and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *Am. J. Pol. Sci.* 52, 705–722. doi: 10.1111/j.1540-5907.2008.00338.x
- Spirling, A. (2023). Why open-source generative AI models are an ethical way forward for science. *Nature* 616, 413. doi: 10.1038/d41586-023-01295-4
- Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., et al. (2020). Neural machine translation: a review of methods, resources, and tools. *AI Open* 1, 5–21. doi: 10.1016/j.aiopen.2020.11.001
- Tolochko, P., and Boomgaarden, H. G. (2019). Determining political text complexity: conceptualizations, measurements, and application. *Int. J. Commun.* 13, 1784–1804.
- Volkens, A., Bara, J., Budge, I., McDonald, M. D., and Klingemann, H.-D. (eds) (2013). *Mapping Policy Preferences from Texts: Statistical Solutions for Manifesto Analysts*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199640041.001.0001
- Washington, A. R. (2023). Semantic and semiotic flows: examining variations and changes of “the N-Words” within an indexical field of dynamic meanings. *Atl. Stud.* 1–28. doi: 10.1080/14788810.2023.2235204. [Epub ahead of print].
- Wyndham, A. (2021). *Inside DeepL: The World’s Fastest-Growing, Most Secretive Machine Translation Company*. Zurich: Slatro AG.
- Zumbach, D., and Bauer, P. C. (2021). *deeplr: Interface to the ‘DeepL’ Translation API: R package version 2.0.0*. Available online at: <https://github.com/zumbov2/deeplr>