Shaping Language Translation Technology: The Influence of Social Forces and Values on Natural Language Processing Applications

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Cory Ooten

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor
Travis Elliott, Department of Engineering and Society

Introduction

Natural Language Processing (NLP) is a field of study that combines computer science, artificial intelligence, and linguistics to enable computers to process and analyze human language. The goal of NLP is to create computer programs that can read, understand, and generate natural language which can be applied in a wide range of fields, including healthcare, education, business, and more (Allen, 2003). One of the most prominent applications of NLP is in the area of language translation, where it has the potential to greatly facilitate communication between individuals who speak different languages (Zong & Hong, 2018). However, the implementation of NLP applications is not a neutral process determined solely by technical factors, but is instead influenced by a range of social and cultural factors. Understanding the interplay between these factors is critical to developing NLP applications that are not only effective, but also ethical and socially responsible. In this paper, we will explore the influence of social forces and values on NLP applications for language translation, drawing on the insights provided by the Social Construction of Technology (SCOT) framework. Specifically, we will analyze how different social groups and actors shape the development and use of NLP for language translation, how NLP is socially constructed, and how the blackboxing of NLP affects its development and use. Ultimately, this paper argues that a deeper understanding of the social dimensions of NLP applications for language translation is necessary for ensuring that these technologies are developed and used in ways that align with broader social goals and values.

Overview of NLP Applications and the SCOT Framework

Applications of Natural Language Processing are designed to enable computers to understand and generate natural language. This is achieved through various techniques, including statistical models, rule-based systems, and deep learning methods like neural networks. These techniques allow computers to recognize patterns in language data and make predictions based on those patterns. Some common NLP applications include information retrieval, sentiment analysis, speech recognition, and machine translation (Allen, 2003). Information retrieval is involves developing search engines that can understand natural language queries, while sentiment analysis can be used to analyze social media posts, customer feedback and other text data to determine the sentiment or emotion expressed by the author. Speech recognition can help transcribe spoken language into text, while machine translation is used to automatically translate text from one language to another (Zong & Hong, 2018).

The Social Construction of Technology (SCOT) framework is a theoretical approach that has been widely used to analyze the relationship between technology and society. Proposed by Hans Klein and Daniel Lee Kleinman (2002), the SCOT framework emphasizes that technology is not a fixed object or objective entity, but is instead shaped by the social and cultural context in which it is developed and used. According to SCOT, technologies are the products of social groups and actors who negotiate and contest over their meanings, uses, and functions (Klein & Kleinman, 2002). At the core of the SCOT framework are several key concepts, including interpretive flexibility, the role of relevant social groups, and the importance of technological artifacts in shaping social relations (Klein & Kleinman, 2002). Interpretive flexibility refers to the idea that technologies can be interpreted and used in multiple ways by different social groups depending on their needs and interests. Relevant social groups are those who have a stake in the development and use of that technology, and ultimately, they have an interest in shaping the outcome of that technology. These components of SCOT demonstrate that technological artifacts are not neutral tools, but instead are filled with social meanings and values that reflect the interests and perspectives of their designers and users. It is essential to analyze these factors when considering technologies, and will be critical for assessing the performance and effectiveness of NLP in language translation.

The Social Construction of NLP Applications for Language Translation

The development and use of NLP applications for language translation are shaped by a range of social groups and actors. These groups have varying levels of influence, and can have more say in how the NLP application is shaped. One of the most important social groups are language communities (Nwokediuko, 2012). Understandably, language communities play a crucial role in shaping an application designed for language translation. These are the people who will use the application to communicate with individuals who speak different language. As such, they have influence when providing feedback on the effectiveness and usability of these applications. For example, Spanish-speaking communities may have different needs and preferences than Chinese-speaking communities when it comes to language translation tools. Some other language communities by be able to even have unique insights into the effectiveness and usability during development of these applications. Language communities can also bring their own linguistic and cultural perspectives to the development of NLP applications (Maxwelll-Smith et al., 2020). These perspectives inform the design of language translation applications and help ensure they are culturally and linguistically appropriate. Bilingual speakers, for instance, can provide valuable feedback on how well an NLP tool captures the nuances of their language, and whether it accurately reflects the way that they speak their language in their daily lives (Maxwelll- Smith et al., 2020). These individuals can give input on the translations of slang or inform developers of words to avoid that are insensitive or have additional meanings. Language communities could also affect which features are prioritized or emphasized during development. A culture that deeply values and respects nature may advocate for the development of a precise translation tool that can accurately convey the meaning of words associated with the natural world (Chi et al., 2023). Incorporating a diverse range of viewpoints can enhance the development of NLP applications, ensuring their cultural and linguistic appropriateness.

Governments and policymakers also play an important role in shaping NLP applications for language translation. They can influence the development and use of these applications through laws and regulations, funding decisions, and public policies. For example, governments may fund research into NLP applications for language translation, or they may require that certain types of documents be translated using specific NLP applications (Iriberri & Leroy, 2007). Additionally, Technology developers and companies are key actors in the development and deployment of NLP applications for language translation. They are responsible for creating the algorithms and software that power these applications. They also have a significant influence on the direction and scope of NLP research. Technology developers and companies are driven by economic interests and may prioritize certain applications based on their potential profitability. However, they can also be influenced by broader social values and goals, such as ethical considerations and social responsibility.

The SCOT framework also highlights the concept of interpretive flexibility, referring to the idea that different actors can have different interpretations of a technology, and this can lead to variations in its development, usage, and impact (Klein & Kleinman, 2002). The social groups that were just discussed demonstrate the concept of interpretive flexibility as seen in SCOT. Some language communities may place a high value of preserving their language and culture and may be resistant to using NLP translation applications (Chang et al., 2019). Others may see these applications as a valuable tool for promoting cross-cultural understanding and be more open to their use. Some actors may use the application for political purposes, such as in a diplomatic context where it is used to facilitate negotiations between countries. For businesses, the focus may be placed on producing fast and accurate translations that allow for efficient communication and transactions (Gilson & Weyns, 2019). In all of these cases, the interpretive flexibility of NLP translation applications demonstrate the ways in which social and cultural factors can shape the development and use of its technology. It also shows the importance of considering the perspectives and values of different actors in the design and implementation of these technologies.

The Blackboxing of NLP

Blackboxing is a term used to describe the process by which complex systems, such as NLP applications, are made to appear as “black boxes” to users, meaning that the inner workings of the system are hidden from view (Choudhary et al., 2022). The intent of blackboxing is typically to protect intellectual property or trade secrets, where companies or individuals may purposefully keep certain aspects of their technology or products hidden from competitors. However, blackboxing can also have negative consequences, such as when it is used to conceal bias or unethical practices within a system. This can prevent users from understanding how decisions are being made or from identifying and addressing potential issues or injustices. In such cases, blackboxing can be seen as a way to avoid accountability as it becomes difficult to identify who or what is responsible for the decisions made by the system. For example, if an NLP application makes an inaccurate translation, it may be unclear who is responsible for the error. This can make it difficult to assign blame and to ensure that appropriate action is taken to address the problem. It can also limit transparency, as users are unable to understand how the system arrives at its decisions. This can lead to mistrust and undermine the user's confidence in the system. Lack of transparency can also limit the ability of researchers and developers to identify and address potential biases or errors in the system (Tsvetkov et al., 2019).

NLP applications are often blackboxed through the use of proprietary algorithms and closed-source software. Language translation software is no different from many other software products developed by companies. Blackboxing is used to protect intellectual property and stop replication of ideas by competitors, but it can create problems. In addition to the previously mentioned accountability and transparency issues, the blackboxing of NLP applications can raise ethical considerations, such as concerns about bias, fairness, and privacy (Yanbo, 2020).

A major ethical concern of nearly all NLP applications is bias. If the algorithms and decision-making processes within an NLP application are not transparent, it can be difficult to determine whether they are biased against certain groups or individuals. For example, if an NLP application used to screen job candidates has been trained on historical data that reflects biases against certain groups, the application may unfairly disadvantage those groups in the screening process. Blackboxing the application can make it difficult to get to the root of this issue (Tsvetkov et al., 2019). Similarly, the blackboxing of NLP applications can also raise concerns related to fairness. If an NLP application is making decisions that affect people's lives, such as decisions related to credit scoring or healthcare, it is important that those decisions are made in a fair and transparent manner. Blackboxing can make it difficult to ensure that the decisions being made are fair and unbiased (Grandvoinnet et al., 2015). There can also be issues related to privacy created by blackboxing. If the data being processed by an NLP application is not transparently and securely handled, it can put people's personal information at risk. This can be particularly concerning in cases where the data being processed is sensitive or personal, such as in healthcare or financial applications (Grandvoinnet et al., 2015). Unfortunately, blackboxing is not an entirely negative process. There are significant benefits of it that were mentioned before, and it can be hard to justify that companies should stop doing so. As such, the issues it can produce are important to understand as they may be here to stay.

Case Studies

NLP language translation applications have become an integral part of our daily lives. These apps have made it easier than ever before to communicate and interact with people from different parts of the world, breaking down language barriers and fostering greater global connectivity. However, these applications are not without their issues. Even the most popular applications have had their share of ethical concerns. Google Translate has been a widely used NLP application for language translation since its inception in 2006. However, it has been criticized for perpetuating gender stereotypes in its translations. This issue formed due to how Google Translate has been "trained", referring to how the algorithm learns from data given to it. Google Translate has been trained on vast amounts of text from the internet, which reflects the biases and stereotypes present in society. In many cases, the text used to train Google Translate contains gender-specific language that reinforces stereotypes about gender roles and professions. For example, a 2016 study (Bolukbasi et al., 2016) found that Google Translate consistently assigned gender to certain professions. For instance, the study found that doctors and engineers were frequently to be male, and nurses and teachers were assigned to be female. Additionally, the study also found that certain adjectives were associated with specific genders as well. "Strong" and "powerful" were translated to male in many languages, while "weak" and "passive" were translated to female (Bolukbasi et al., 2016).

To address these issues, Google has taken steps to improve the gender sensitivity of its translations. In 2018, it introduced gender-specific translations for a limited set of languages, allowing users to choose between masculine and feminine translations for certain phrases. It also began using machine learning to detect and reduce gender bias in translations, and worked with linguists to identify and address gender-specific issues in different languages (Fitria, 2021).

However, Google is not the only company to experience issues with their NLP applications. In 2016, a popular NLP application called “Tay” was launched by Microsoft. Tay was an experimental chatbot designed to learn and interact with users via Twitter, with the aim of improving its conversational skills over time. However, within 24 hours of its launch, Tay began to produce racist and offensive content in its responses to users (Victor, 2016). Tay’s responses included a range of offensive and racist comments, including references to Hitler, support for genocide, and racist slurs. These extreme responses were not values that Microsoft designed Tay to have. Tay’s machine learning algorithms were designed to learn from interactions with other Twitter users, and to adjust its responses based on the language and content of those interactions. However, as users began to interact with Tay, they intentionally fed it with offensive content and Tay learned from those users.

The developers of Tay had not anticipated the extent to which users would intentionally try to manipulate the chatbot, and as a result, Tay's responses were not properly monitored or controlled (Wolf et al., 2017). As a result, Tay's learning algorithms were influenced by the harmful biases and prejudices present in the interactions it had with users, ultimately leading to its inappropriate responses. This case highlights the importance of careful monitoring and control in the development of NLP language applications. Machine learning algorithms can be powerful tools for improving the accuracy and effectiveness of these applications, but they are also vulnerable to the biases and prejudices of the people who interact with them (Yanbo, 2020). As such, it is essential to consider the potential risks and challenges involved in the development of these tools, and to take steps to mitigate these risks as much as possible.

Conclusion

The study of natural language processing applications for language translation through the lens of the social construction of technology theory provides valuable insights into the social, cultural, and political factors that shape the development and use of these applications. We have explored the social construction of NLP applications, which are not neutral tools but rather products of social groups and actors that bring their own interests, values, and perspectives to the table. Additionally, we have discussed the blackboxing of NLP, which creates significant challenges for ensuring accountability, transparency, and ethical considerations in the development and use of these applications. Through the studies of Google Translate and Microsoft's Tay, we have seen the real-world implications of the issues surrounding NLP applications for language translation. As technology continues to advance and become more integrated into our lives, it is essential that we critically examine and understand the ways in which these tools are designed, developed, and deployed. Only then can we work towards ensuring that NLP applications for language translation are developed and used in a socially responsible and ethical manner.

References

Allen, J. F. (2003). Natural Language Processing. In *Encyclopedia of Computer Science* (pp. 1218–1222). John Wiley and Sons Ltd.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings* (arXiv:1607.06520). arXiv. https://doi.org/10.48550/arXiv.1607.06520

Chang, K.-W., Prabhakaran, V., & Ordonez, V. (2019, November). *Bias and Fairness in Natural Language Processing*. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): Tutorial Abstracts. https://aclanthology.org/D19-2004

Chi, Y., Kim, J. H., & Sun, S. (2023). Korean Language NLP Model Based Emotional Analysis of LGBTQ Social Media Communities. *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 1–5. https://doi.org/10.1109/IMCOM56909.2023.10035659

Choudhary, S., Chatterjee, N., & Saha, S. K. (2022). *Interpretation of Black Box NLP Models: A Survey* (arXiv:2203.17081). arXiv. https://doi.org/10.48550/arXiv.2203.17081

Fitria, T. N. (2021). *Gender Bias in Translation Using Google Translate: Problems and Solution* (SSRN Scholarly Paper No. 3847487). https://papers.ssrn.com/abstract=3847487

Gilson, F., & Weyns, D. (2019). When Natural Language Processing Jumps into Collaborative Software Engineering. *2019 IEEE International Conference on Software Architecture Companion (ICSA-C)*, 238–241. https://doi.org/10.1109/ICSA-C.2019.00049

Grandvoinnet, H., Aslam, G., & Raha, S. (2015). *Opening the Black Box: The Contextual Drivers of Social Accountability*. World Bank Publications.

Iriberri, A., & Leroy, G. (2007). Natural Language Processing and e-Government: Extracting Reusable Crime Report Information. *2007 IEEE International Conference on Information Reuse and Integration*, 221–226. https://doi.org/10.1109/IRI.2007.4296624

Klein, H., & Kleinman, D. (2002). The Social Construction of Technology: Structural Considerations. *Science Technology & Human Values - SCI TECHNOL HUM VAL*, *27*, 28–52. https://doi.org/10.1177/016224390202700102

Maxwelll-Smith, Z., González Ochoa, S., Foley, B., & Suominen, H. (2020). Applications of Natural Language Processing in Bilingual Language Teaching: An Indonesian-English Case Study. *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 124–134. https://doi.org/10.18653/v1/2020.bea-1.12

Nwokediuko, A. A. (2012). Advancing e-learning in African native communities: The language factor. *International Conference on Education and E-Learning Innovations*, 1–4. https://doi.org/10.1109/ICEELI.2012.6360594

Tsvetkov, Y., Prabhakaran, V., & Voigt, R. (2019). Socially Responsible Natural Language Processing. *Companion Proceedings of The 2019 World Wide Web Conference*, 1326. https://doi.org/10.1145/3308558.3320097

Victor, D. (2016, March 24). Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk. *The New York Times*. https://www.nytimes.com/2016/03/25/technology/microsoft-created-a-twitter-bot-to-learn-from-users-it-quickly-became-a-racist-jerk.html

Wolf, M. J., Miller, K. W., & Grodzinsky, F. S. (2017). Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications. *The ORBIT Journal*, *1*(2), 1–12. https://doi.org/10.29297/orbit.v1i2.49

Yanbo, Z. (2020). Implicit Bias or Explicit Bias: An Analysis Based on Natural Language Processing. *2020 International Conference on Computing and Data Science (CDS)*, 52–55. https://doi.org/10.1109/CDS49703.2020.00018

Zong, Z., & Hong, C. (2018). On Application of Natural Language Processing in Machine Translation. *2018 3rd International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, 506–510. https://doi.org/10.1109/ICMCCE.2018.00112