Health Care Misinformation: An artificial intelligence challenge for low-resource languages

Sarah Luger[1], Martina Anto-Ocrah[2], Tapo Allahsera[3], Christopher M. Homan[3], Marcos Zampieri[3], Michael Leventhal[4]

[1]Orange Silicon Valley, San Francisco, CA, USA
[2]University of Rochester Medical Center, Rochester, NY, USA
[3]Rochester Institute of Technology, Rochester, NY, USA
[4]Centre National de l'Education en Robotique et en Intelligence Artificielle (RobotsMali), Bamako, Mali

sarah.luger@orange.com | martina_anto-ocrah@urmc.rochester.edu | (aat3261 | cmhvcs | mazgla)@rit.edu
mleventhal@robotsmali.org

Abstract

In this paper, we motivate using state-of-the-art artificial intelligence technologies to address challenges presented by low-resource languages. We also reflect on both the importance and priorities of AI research with respect to the less wealthy economies of the world. We explore the contributions of colonialism to language (in)accessibility and public health misinformation during the Covid-19 pandemic in the African region. Using the West African country of Mali as a case study, we discuss the historic contribution of colonial educational systems to the creation of disenfranchised populations. These populations are left with limited access to important medical information that can mean life or death in the current Covid-19 pandemic. We propose a humans-in-the-loop neural machine translation, (NMT), solution to medical information translation. In our solution, the state-of-the-art NMT approach is applied to the low-resource language Bambara which is spoken by a majority of the Malian people. By implementing a crowdsourced Bambara language data collection and translation component in this machine learning problem, we engage the local Malians. The aim of this project is to address the lack of Bambara language resources and leverage current best practice in order to undo some of the artefacts of colonialism. We describe the unique challenges and research issues raised by this novel application of AI technology.

Background

AI research can contribute to the diminution of global inequities borne of the colonial period. However, the research community, with notable exceptions, fails either to recognize this opportunity or to be interested in it because research priorities reflect the same mindset that produced colonialism. Based on well-publicized instances of bias in AI systems over the past several years (Buolamwini and Gebru 2018; Angwin et al. 2016), the priority for current "AI for Social Good" or fairness, accountability, and transparency research is to address problems that reflect Western challenges with institutional racism and sexism. This Western-centric approach, while nascent, misses opportunities to increase financial, educational, and health well-being elsewhere. There is tremendous opportunity, especially in machine translation, (MT), of medical information, to increase digital inclusion and health for low-resource language speakers.

Andrew Ng, through his deeplearning.ai newsletter, ran a survey asking what the AI community should focus on in order to promote social good. The authors of this paper believe that a top focus should be to solve problems in developing countries where it could have an enormous impact and to help create expertise in the developing world. This big impact and increased expertise means the people that live in the developing world can control their own technology and their own destiny.

In 2018, the McKinsey Global Institute published research outlining the financial benefits of corporate and national AI investment. One of their top four analyses was that

> [the] adoption of AI could widen gaps between countries, companies, and workers...AI leaders (mostly developed economies) could capture an additional 20 to 25 percent in economic benefits compared with today, while emerging economies may capture only half their upside (Bughin et al. 2018).

Another of the top four analyses was:

> The pace of AI adoption and...how countries choose to embrace these technologies (or not) will likely impact the extent to which their businesses, economies, and societies can benefit. The race is already on among companies and countries. In all cases, there are trade-offs that need to be understood and managed appropriately in order to capture the potential of AI for the world economy (Bughin et al. 2018).

At this juncture, artificial intelligence is a field that faces both vast promise and daunting peril. We seek to raise

AAAI Fall 2020 Symposium on AI for Social Good.

awareness of the challenges faced by communities isolated by a lack of language resources, especially digital ones. In addition, we present a repeatable use case for low-resource languages: using neural machine translation with humans-in-the-loop to improve global access to health care information.

As a team working on AI with the full participation of an African research team, on a project for Africans, we encounter the notion of "social good" frequently, as well as the intrinsically related concepts of "fairness," "accountability," and "transparency." We have formed the view that it almost always turns around the problems and perspective as perceived and as defined in the societies of plenty. When we (some of the authors), as Africans, look from any side of the political spectrum at the debate in the developed countries, we do not sense that there is real conviction that our fates are interlinked on a global level.

Wealthier nations may feel that international institutions, which they fund, are addressing global needs. Many Africans understand that the primary function of these institutions is to keep African suffering out of the developed world. Wealthier nations may point to foreign aid as their contribution to alleviating that suffering. Many Africans see that foreign aid is always tied narrowly to the priorities defined by the donors and not the people purportedly aided, that the great bulk of the money goes back into the donor country through salaries paid to consultants and goods purchased in the donor country, and that the overall effect is to suppress the development of local industry and expertise.

Preliminary findings from our work shows that what many Africans would love to have instead is access to the same resources that many people in the wealthier economies and past colonial powers have to educate themselves, to start businesses, and to create opportunities and solutions to the problems that they face. There are many systematic ways that these countries historically have labored to deny such access to ex-colonial countries and there continue, to this day, to be numerous systematic ways that they continue to do this.

Challenge

This paper begins with background information on the suppression of native-language education in the era of colonialism as an paradigmatic historical example of a widespread, long-term policy to deny Africans the access to resources to develop their own intellectual capacity and the capacity to solve problems relevant to them. We argue that the continuity of the colonial mindset is reflected in the fact that, today, only minuscule resources exist for Africans to learn AI, that Africans are systematically, en masse, denied access to resources to learn AI and participate in AI communities that exist only outside of Africa, and that "AI for Social Good" has not even considered a problem as basic as applying natural language processing, (NLP), to the languages that Africans speak.

We present a case study covering the impact of this inattention and denial of resources on health care information in Africa as the Covid-19 epidemic swept the world. We use our own efforts to study the problems of developing NLP for an under-resourced African language, Bambara, as a generalizable use case. Exploring Bambara MT illustrates that the AI challenges and research problems aiming to do social good must also reflect the priorities of the inhabitants of financially under-resourced countries.

Colonialism and its Legacy

To gain a sense of the significance of misinformation to public health crises, consider the situation in Mali in May, 2020. As the death toll at that time had reached over 300,000 people globally, and news of the increasing Covid-19 mortality rates dominated media outlets, the United Nations announced the "Verified" initiative (Kreider 2020) to fight Covid-19 misinformation. And yet it was only in the following month that they began donating medical relief to Mali in order to support an integrated and quick response to the Covid-19 crisis (for the Coordination of Humanitarian Affairs 2020).

French and British colonial language perspectives

Since the colonial era, language has served to disenfranchise African populations. However, different approaches to colonization by France and Britain led to very different outcomes in this regard. In order to save costs and provide the appearance of a moral justification for colonialism, the British relied on missionaries to manage education in their colonies. This approach was inherently decentralized, with individual missions having great liberty in how they taught. This allowed them to provide most instruction in the local vernacular, and teach English as a second language as a specific topic.

France, by contrast, used language to drive assimilation and effectively "turn" Africans into French people (Cogneau and Moradi 2014; Benavot and Riddle 1988; Garnier and Schafer 2006). Schools required government certification, the hiring of government-certified teachers, and adherence to a government-sanctioned curriculum. All instruction was in French only. The colonial state was the primary educator, and only those who could navigate the administrative and cost barriers received an education.

These divergent approaches led to significant disparities in school enrollment and literacy levels in the colonies, with higher school enrollments and literacy in the less centralized British-format system, compared to the more centralized French system (Cogneau and Moradi 2014; Benavot and Riddle 1988; Garnier and Schafer 2006).

Modern-day ramifications of colonial language on the Covid-19 crisis in Mali

In Mali, which was colonized by France for 68 years, French remains the official language. Yet only 20% of the population have mastered it, due to the high costs of and barriers to educational resources (Mingat and Suchaut 2000; ArcGIS 2020). Most Malians are multilingual, and the majority of them speak Bambara, the primary language of the predominant ethnic group (Mingat and Suchaut 2000). Due to a paucity of information about Covid-19 in Bambara, those 15.2 million Malians with fluency in Bambara but not

French have limited access to critical public health information, such as viral transmission modes, use of personal protective equipment, movement restrictions, quarantine measures, and social distancing protocols. Absent the capacity to widely disseminate crucial, novel information, efforts to combat Covid-19 in some of the most vulnerable and disenfranchised Malian communities continues to be challenging.

Using AI to improve health information

In this section we present our strategy to improve Bambara language resources. We begin with leveraging emergent neural machine translation technology which relies on aligning corresponding text from Bambara and French. Then, we describe our preliminary study which uses a relatively small amount of data and helps identify the challenges of human translation for Bambara and similar, primarily spoken, languages. Finally, we discuss the importance of crowdsourcing and the development of our neural machine translation system.

Proposed Solution

Text alignment is a process that creates a correspondence from a ground truth translation to that of a novel generated translation. In situations like this with low resource languages, alignment begins by using a trained Bambara to French translator on a data set of Bambara to French sentences to create a loose correlation between the sets. From there, an automated aligner processes the translated French sentences and the ground truth French to create an "alignment".

Word alignment models (Och and Ney 2004) are very important in neural and statistical MT pipelines. Poor alignment performance tends to lead to poor MT performance. Several studies have investigated the relation between high-quality word alignment and MT quality in terms of automatic metrics such as BLEU scores (Fraser and Marcu 2007). Obtaining high quality word alignment depends on the availability of suitable (often large) parallel corpora which is a known challenge for low-resource languages like Bambara. There have been studies proposing methods to improve word alignment models for low resource language pairs (Xiang, Deng, and Zhou 2010; McCoy and Frank 2018) including the use of a resource-richer pivot language to improve word alignment between a low resource pair (triangulation) (Levinboim and Chiang 2015), however, to the best of our knowledge, there have been no studies addressing Bambara specifically.

As noted, building Bambara language capacity in Mali via MT requires constructing Bambara-language information from source data in another language (ACALAN 2020). Quickly scaling MT technology however depends on sufficient amounts of translated text from source to target language to train the translation system before it can achieve state-of-the-art levels. Bambara lacks such training data and has been considered (from the perspective of MT training data) a low-resource language (Wu et al. 2016).

Thus, MT technology that uses a humans-in-the-loop approach can engage local Malians to bridge the language divide. Using crowdsourcing platforms, Malians can be resourced to translate small amounts of Bambara to French (and vice versa). This crowdsourcing process can create sufficient training data necessary for implementing MT technology (Wu et al. 2016; Leventhal et al. 2020). Crowdsourcing begins the digital data development cycle aimed at transitioning Bambara out of the low-resource language category. Highly digitally-resourced languages leverage sufficient data to improve the quality of their automated translations. This transition would also reduce unnecessary burdens placed on local governments who are plagued with the devastating Covid-19 pandemic, whilst still reeling from the effects of colonialism.

There have been many attempts to use machine translation for Covid-19 response (Way et al. 2020; TAUS 2020; without borders 2020; Project 2020), but only the last two of these, Translators without Borders (without borders 2020) and The Endangered Languages Project (Project 2020) consider African languages. We see these efforts as motivation for bottom-up solutions through crowdsourcing so that their same success in MT modeling can be achieved for Bambara. Broadly, our goal is to use Bambara as a test case for modeling best practice for future initiatives in low resource language data collection, crowdsourced labor training and annotation, and high-quality NMT model building.

Preliminary study

We undertook a preliminary study of NMT, collecting data and creating a model to translate between Bambara and French and English. The goal of this work was not only to elucidate the challenges of NLP for this particular language and, in general, for under-resourced languages, but also to gather data for the preparation of a full-scale attack on the problem. This work is described in more detail in (Luger, Homan, and Tapo 2020; Leventhal et al. 2020).

Data Collection and Preparation

The data for our initial study is a dictionary dataset from *SIL Mali*[1] with examples of sentences used to demonstrate word usage in Spanish, French, English, and Bambara; and a tri-lingual health guide titled “Where there is no doctor.[2]”

Data preparation, including alignment, proved to be about 60% of the overall time spent in person-hours on the experiment and required on-the-ground organization and recruitment of skilled volunteers in Mali.

Most of the dictionary examples of expressions in Bambara are formatted as dictionary entries followed by their translations in French and in English. Most of these are single sentences, so there is sentence-to-sentence alignment in the majority of cases. However, there remains a sufficient number of exceptions to render automated pairing impossible. Part of the problem lies in the unique linguistic and cultural elements of the bambaraphone environment; it is often not possible to meaningfully translate an expression in Bambara without giving an explanation of the context.

[1] https://www.sil-mali.org/en/content/introducing-sil-mali

[2] https://gafe.dokotoro.org/

The medical health guide is aligned by chapters, each of which is roughly aligned by paragraphs. But at the paragraph level there are too many exceptions for automated pairing to be feasible. Further, at the sentence level many of the bambaraphone-specific problems found in the dictionary dataset are present here, particularly in explanations of concepts that can be succinctly expressed in English or French but for which Bambara lacks terminology and the bambaraphone environment lacks an equivalent physical or cultural context.

Both datasets required manual alignment by individuals fluent in written Bambara and either French or English, and able to exercise expert-level judgment on linguistic and, occasionally, medical questions. Access to such human expertise was a major factor limiting the quantity of data we were able to align. We implemented a software alignment tool to manually align sentences and to save those sentence pairs that a human editor considered properly aligned. In separate tasks, four annotators with a middle school level understanding of Bambara performed alignment on French-Bambara and English-Bambara sentence pairs using the tool.

The final dataset contains 2,146 parallel sentences of Bambara-French and 2,158 parallel sentences of Bambara-English–a tiny amount of data for NMT compared to massive state-of-the-art models that are trained on millions of sentences (Arivazhagan et al. 2019).

Thus, our NMT is a transformer (Vaswani et al. 2017) of appropriate size for a relatively smaller training dataset (van Biljon, Pretorius, and Kreutzer 2020). It has six layers with four attention heads for encoder and decoder, the transformer layer has a size of 1024, and the hidden layer size 256, the embeddings have 256 units. Embeddings and vocabularies are not shared across languages, but the softmax layer weights are tied to the output embedding weights. The model is implemented with the Joey NMT framework (Kreutzer, Bastings, and Riezler 2019) based on PyTorch (Paszke et al. 2019).

Training runs for 120 epochs in batches of 1024 tokens each. The ADAM optimizer (Kingma and Ba 2014) is used with a constant learning rate of 0.0004 to update model weights. This setting was found to be best to tune for highest BLEU (Papineni et al. 2002), compared to decaying or warmup-cooldown learning rate scheduling. For regularization, we experimented with dropout and label smoothing. The best values were 0.1 for dropout and 0.2 for label smoothing across the board. For inference, beam search with width of 5 is used. The remaining hyperparameters are documented in the Joey NMT configuration files.

Neural Machine Translation Results

Translation results were evaluated both automatically and with human evaluators. We obtained BLEU scores of approximately 20 for our best model. BLEU or "bilingual evaluation understudy" is a system of measuring automated machine translation's text output with high scores being closest to those of a professional human translator (Papineni et al. 2002).

Two human evaluators, native speakers of Bambara and self-assessed to be fluent in English and French, evaluated a random sampling of 41 translations of Bambara, 21 into English and 20 into French. The evaluators did not collaborate with each other. The evaluators were asked to assess several aspects of the translations, including identifying specific parts that were well or poorly translated. Finally, the evaluators were asked to identify those translations that succeeded in conveying most of the meaning of the Bambara source, and to assign them a quality score. Of these 41 sentences, one evaluator classified 8 sentences as nearly perfect or very good while the second gave 17 this rank. All 8 of the first evaluator's translations were selected by the second. The Cohen Kappa score of the pair is 0.5141 indicating moderate agreement (Viera and Garrett 2005).

Our analysis suggests that we did not provide sufficient guidance as to what constitutes an acceptable translation to our human Bambara evaluators. Further, one evaluator was simply more lenient than the other in what they deemed was acceptable for meeting the subjective label of "nearly perfect or very good translation". Moreover, we had difficulty formulating translation criteria due to limited experience with human translation of Bambara, in addition to the *ab initio* nature of this experiment with machine translation of Bambara. Moving forward, our results will inform the development of more rigorous criteria in future experiments.

Conclusion

Our study constitutes the first attempt of modeling automatic translation for the low-resource language of Bambara. We identified challenges for future work, such as the development of alignment tools for small-scale datasets, the need for a general domain evaluation set, and better training of human translation evaluators. The current limitation of processing written text as input might furthermore benefit from the integration of spoken resources through speech recognition or speech translation, since Bambara is primarily spoken and the lack of standardization in writing complicates the creation of clean reference sets and consistent evaluation.

Future Work

Moving forward we would like to take advantage of the human-in-the-loop approach described here to create more resources to improve word alignment and MT systems for low-resource languages in general and Bambara in particular. Another avenue we would like to explore is the use of monolingual data. The health care domain is rich in resources for English (e.g. UMLS [3], SNOMED [4], NCBO's BioPortal[5]) and such monolingual data can be used to improve the performance of MT systems on the English side of the English–Bambara translation pair (Burlot and Yvon 2019). Finally, the use of term banks, either manually or automatically compiled, is another under-explored avenue for low-resource languages (Haque, Penkale, and Way 2014) which we believe can be particularly helpful for technical domains such as medicine and health care.

[3] https://www.nlm.nih.gov/research/umls/index.html

[4] http://www.snomed.org/

[5] https://bioportal.bioontology.org/

In addition, we have made data sets, including aligned and annotated French and Bambara sentence pairs available to the machine translation and AI for Good community: *Bambara Data Repository*[6]. Please reach out to us regarding these low-resource language resources as we are attempting to make as much of our research as possible available to the community.

Acknowledgments

We would like to thank Julia Kreutzer, Arthur Nagashima, the Masakhane machine translation for Africa community, and *SIL Mali*[7]. Our work could not have been possible without your valuable insight and contributions to ongoing progress in this field. Earlier versions of this work are described in (Luger, Homan, and Tapo 2020; Leventhal et al. 2020).

References

ACALAN. 2020. African academy of languages, african union. historical background. https://acalan-au.org/aboutus.php.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine bias. *ProPublica, May* 23:2016.

ArcGIS. 2020. Mali languages. https://www.arcgis.com/home/item.html?id=b5b4f736b5714f32b12a0322e5405734.

Arivazhagan, N.; Bapna, A.; Firat, O.; Lepikhin, D.; Johnson, M.; Krikun, M.; Chen, M. X.; Cao, Y.; Foster, G.; Cherry, C.; Macherey, W.; Chen, Z.; and Wu, Y. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges.

Benavot, A., and Riddle, P. 1988. The expansion of primary education, 1870-1940: Trends and issues. *Sociology of Education* 61:191.

Bughin, J.; Seong, J.; Manyika, J.; Chui, M.; and Joshi, R. 2018. Mckinsey global institute notes from the ai frontier: Modeling the impact of ai on the world economy. *McKinsey Global Institute, September* 1:64.

Buolamwini, J., and Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Friedler, S. A., and Wilson, C., eds., *Conference on Fairness, Accountability and Transparency, FAT 2018, 23-24 February 2018, New York, NY, USA*, volume 81 of *Proceedings of Machine Learning Research*, 77–91. PMLR.

Burlot, F., and Yvon, F. 2019. Using monolingual data in neural machine translation: a systematic study. *arXiv preprint arXiv:1903.11437*.

Cogneau, D., and Moradi, A. 2014. British and french educational legacies in africa. https://voxeu.org/article/british-and-french-educational-legacies-africa.

for the Coordination of Humanitarian Affairs, U. N. O. 2020. Support for mali's covid-19 response plan: The united nations hands over 48 tons of medical supplies and equipment. https://reliefweb.int/report/mali/support-malis-COVID-19-response-plan-united-nations-hands-over-48-tons-medical-supplies.

Fraser, A., and Marcu, D. 2007. Measuring word alignment quality for statistical machine translation. *Computational Linguistics* 33(3):293–303.

Garnier, M., and Schafer, M. 2006. Educational model and expansion of enrollments in sub-saharan africa. *Sociology of Education - SOCIOL EDUC* 79:153–176.

Haque, R.; Penkale, S.; and Way, A. 2014. Bilingual termbank creation via log-likelihood comparison and phrase-based statistical machine translation. In *Proceedings of the 4th International Workshop on Computational Terminology (Computerm)*, 42–51.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kreider, K. 2020. Why the western system of covid-19 response won't work in africa. https://theowp.org/reports/why-the-western-system-of-COVID-19-response-wont-work-in-africa/.

Kreutzer, J.; Bastings, J.; and Riezler, S. 2019. Joey NMT: A minimalist NMT toolkit for novices. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, 109–114. Hong Kong, China: Association for Computational Linguistics.

Leventhal, M.; Tapo, A.; Luger, S.; Zampieri, M.; and Homan, C. M. 2020. Assessing human translations from french to bambara for machine learning: a pilot study. *arXiv preprint arXiv:2004.00068*.

Levinboim, T., and Chiang, D. 2015. Multi-task word alignment triangulation for low-resource languages. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1221–1226.

Luger, S.; Homan, C. M.; and Tapo, A. 2020. Towards a crowdsourcing platform for low resource languages: A collectivist approach. *AAAI: Human Computation (HCOMP)*.

McCoy, R. T., and Frank, R. 2018. Phonologically informed edit distance algorithms for word alignment with low-resource languages. *Proceedings of the Society for Computation in Linguistics* 1(1):102–112.

Mingat, A., and Suchaut, B. 2000. Les systèmes éducatifs africains. une analyse économique comparative. https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2215667.

Och, F. J., and Ney, H. 2004. The alignment template approach to statistical machine translation. *Computational linguistics* 30(4):417–449.

Papineni, K.; Roukos, S.; Ward, T.; and jing Zhu, W. 2002. Bleu: a method for automatic evaluation of machine translation. 311–318.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga,

[6]https://github.com/israaar/mt_bambara_data

[7]https://www.sil-mali.org/en/content/introducing-sil-mali

L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc. 8026–8037.

Project, E. L. 2020. Covid-19 information in indigenous, endangered, and under-resourced languages. https://endangeredlanguagesproject.github.io/COVID-19/.

TAUS. 2020. Corona corpus. https://md.taus.net/corona.

van Biljon, E.; Pretorius, A.; and Kreutzer, J. 2020. On optimal transformer depth for low-resource language translation. *"AfricaNLP" Workshop at the 8th International Conference on Learning Representations*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Viera, A., and Garrett, J. 2005. Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5):360–3.

Way, A.; Haque, R.; Xie, G.; Gaspari, F.; Popovic, M.; and Poncelas, A. 2020. Facilitating access to multilingual covid-19 information via neural machine translation. *arXiv preprint arXiv:2005.00283*.

without borders, T. 2020. Twb glossary for covid-19. https://translatorswithoutborders.org/twb-glossary-for-covid-19/.

Wu, Y.; Schuster, M.; Chen, Z.; Le, Q. V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; Klingner, J.; Shah, A.; Johnson, M.; Liu, X.; Łukasz Kaiser; Gouws, S.; Kato, Y.; Kudo, T.; Kazawa, H.; Stevens, K.; Kurian, G.; Patil, N.; Wang, W.; Young, C.; Smith, J.; Riesa, J.; Rudnick, A.; Vinyals, O.; Corrado, G.; Hughes, M.; and Dean, J. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation.

Xiang, B.; Deng, Y.; and Zhou, B. 2010. Diversify and combine: Improving word alignment for machine translation on low-resource languages. In *Proceedings of the ACL 2010 Conference Short Papers*, 22–26.