A Comparison of Different NMT Approaches to Low-Resource Dutch-Albanian Machine Translation

Arbnor Rama rama.arbnor@gmail.com
Eva Vanmassenhove e.o.j.vanmassenhove@tilburguniversity.edu
Department of CSAI, Tilburg University, Tilburg, The Netherlands

Abstract

Low-resource languages can be understood as languages that are more scarce, less studied, less privileged, less commonly taught and for which there are less resources available (Singh, 2008; Cieri et al., 2016; Magueresse et al., 2020). Natural Language Processing (NLP) research and technology mainly focuses on those languages for which there are large data sets available. To illustrate differences in data availability: there are 6 million Wikipedia articles available for English, 2 million for Dutch, and merely 82 thousand for Albanian. The scarce data issue becomes increasingly apparent when large parallel data sets are required for applications such as Neural Machine Translation (NMT). In this work, we investigate to what extent translation between Albanian (SQ) and Dutch (NL) is possible comparing a one-to-one (SQ↔AL) model, a low-resource pivot-based approach (English (EN) as pivot) and a zero-shot translation (ZST) (Johnson et al., 2016; Mattoni et al., 2017) system. From our experiments, it results that the EN-pivot-model outperforms both the direct one-to-one and the ZST model. Since often, small amounts of parallel data are available for low-resource languages or settings, experiments were conducted using small sets of parallel NL↔SQ data. The ZST appeared to be the worst performing models. Even when the available parallel data (NL↔SQ) was added, i.e. in a few-shot setting (FST), it remained the worst performing system according to the automatic (BLEU and TER) and human evaluation.

Keywords: *Machine Translation (MT), Neural Machine Translation (NMT), zero-shot MT, pivot-based translation, Dutch-Albanian, NL–SQ, low-resource MT*

1 Introduction

There are more than 7000 languages worldwide, with over 40% of the languages being endangered with less than 1000 speakers. On the other hand, roughly 35% of the world population, close to 3 billion people, account for only 3 languages: English, Mandarin Chinese and Hindi (Eberhard et al., 2021). This language division results in an inequality in literary resources available per language. Languages with a significant amount of literature and speakers are known as high-resource languages in the field of Natural Language Processing (NLP), whilst languages that lack these resources are known as low-resource languages. Mattoni et al. (2017, p.2) define low-resource languages as "languages that have a low population density, are under-taught or have limited written resources or are endangered". These languages are, therefore, not properly represented through literary media, resulting in an insufficient amount of training data availability. One such low-resource language is Albanian.

Albanian is spoken by approximately 8 million people in the world, the majority of which reside in Albania and Kosovo where the language is native to (Kallulli, 2011; Pustina, 2016). The language, however, is not limited to Albania and Kosovo, but extends further into other parts of the Balkans, such as Montenegro, Macedonia and Italy as well as Switzerland, places where Albanian is recognized as a minority language (Kallulli, 2011; Prifti, 2008; Pustina, 2016). Both Prifti (2008) and Pustina (2016) discuss the impact of the Ottoman rule on Albanian literature, a period during which publications in Albanian were forbidden, resulting in little to no development of the written culture. As discussed by Prifti (2008, p.29), this led to a limited number of resources in Albanian, despite the vast number of speakers, books in Albanian were not commonplace until the late 19th century. The impact of which, bred a limited amount of Albanian sources adequately translated into other languages.

A lack of readily available Albanian training data makes developing Statistical MT (SMT) and NMT methods difficult, as these methods require significant amounts of parallel data between language pairs in order to create useful MT systems (Tapo et al., 2020). Additionally, parallel corpora are often times domain-specific, leading to poor performance when deploying MT models for translating material outside of the trained domain (Koehn and Knowles, 2017).

Access to knowledge is a key driver for developing countries to progress in terms of educational, scientific, and societal advancement (Psacharopoulos and Woodhall, 1993). As such, creating opportunities to acquire general knowledge in languages native to developing countries could accelerate the development of their population. One of the most commonly known contributors to online open-access knowledge is Wikipedia (Teplitskiy et al., 2017). However, in terms of accessibility there is a significant lack of articles in non-major languages. For example, there are over 6 million English articles and more than 2 million articles in Dutch, while articles written in Albanian only account for approximately 82 thousand articles.[1]

While the access to knowledge can depend on multiple factors such as, the ability to read and understand English, the access to a stable internet connection, the lack of Albanian training data for NMT models suggest that online literary resources in the language are scarce.[2] Being able to automatically translate Wikipedia articles to a low-resource language offers open-access knowledge to a wider array of people, while allowing these users to improve on the automatically generated translations. Consequently, these improvements can be propagated back to the NMT model, which can help translate future articles more accurately.

In this work, we compare a one-to-one NMT model and two low-resource NMT approaches to translation from Dutch (NL) to Albanian (SQ), a low-resource language pair. By automatically and manually evaluating the translations, we aim to provide insights into how accurately NL↔SQ models can translate. We furthermore explore how the addition of direct parallel NL↔SQ data affects the performance of the ZST model, since often small amounts of parallel data are available. The main research questions can be formulated as follows: (a) *"To what extent are low-resource direct one-to-one NL↔SQ, pivot-based and zero/few-shot NMT models able to accurately translate and how do they compare?"* and (b) *"How does adding parallel NL↔SQ data affect the performance of the ZST model?"*. The performance of the models is evaluated and compared using automatic metrics (BLEU and TER) as well by providing a more detailed human evaluation of 100 random sentences for all models evaluated by three native Albanian speakers.

2 Related Work

SMT and NMT require a significant amount of parallel data in order to produce accurate and fluent translations (Cheng et al., 2017). Advancements in hardware technologies, data augmen-

[1] https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_language_group

[2] https://opus.nlpl.eu/

tation techniques, and deep neural networks, have led to the development of methods capable of translating low-resource languages, subsequently circumventing the need for copious amounts of parallel data (Tapo et al., 2020). As a result, low-resource MT models that use additional languages - also known as a pivot languages - to bypass parallel data between the source and target language have been introduced (Johnson et al., 2016; Ha et al., 2016; Tapo et al., 2020; Liu et al., 2018; Cheng et al., 2017). Traditionally, in the case of a lack of parallel resources, a translation pipeline would be constructed using an intermediate, high-resource pivot language. The pivot-based approach was widely used in the SMT method due to its "simplicity, effectiveness and minimum requirement of multilingual data" (Cheng et al., 2017, p. 3974). The challenge for NMT then, is the lack of large-scale parallel corpora available in order to create better translations.

Johnson et al. (2016) compare how implicit bridging functions in contrast to explicit bridging, for the sake of simplicity, implicit bridging will be referred to as ZST NMT and explicit bridging as pivot MT. Johnson et al. (2016) were the pioneers in showcasing the possibility of a ZST NMT without the use of a(n) (explicit) pivot language. The difference between implicit and explicit bridging is as follows, implicit language bridging allows a system to translate from a source to a target language without having prior training for a specific language pair (Johnson et al., 2016, p. 341). Whereas explicit language bridging requires an extra step where a source language is translated into a pivot language and then from the pivot is translated into the target language (Johnson et al., 2016). Some disadvantages of pivot MT are important to note, namely, a higher total translation time, and the potential for quality loss due to the translation to and from an intermediate language. Further, Johnson et al. (2016) use related languages to investigate the different types of multilingual NMTs, where this paper uses one pair of related languages and an unrelated language – Dutch and English classified as West Germanic languages and Albanian an Indo-European language yet classified as its own subdivision. For our experiments we use Transformers rather than Recurrent Neural Networks (RNN) (Johnson et al., 2016). As in Johnson et al. (2016), we use an additional token that displays the language of origin. A method which resembles Lakew et al. (2018)'s "language flag", where in the pre-processing step a token is embedded into the model so as to identify the target language a source is paired with.

3 Experimental Setup

3.1 Datasets

Parallel data for SQ↔EN, NL↔EN, and SQ↔NL is available in the OpenSubtitles 2018 corpus (Lison and Tiedemann, 2016) which contains movie and TV subtitles for 62 languages total.

Subtitles, from a linguistic perspective, are often referred to as "conversational domain" (Lison and Tiedemann, 2016; Lison et al., 2018). Lison et al. (2018) state that parallel subtitle corpora are used for a variety of NLP tasks, including translation research, conversation models and exploring properties of colloquial language.

Table 1, shows the amount of data files relating to each individual language available. It is important to note that between the OpenSubtitles 2016 and OpenSubtitles 2018 the amount of data increased by more than 25% for both English and Dutch subtitles (Lison et al., 2018). Where Albanian files saw an increase of less than 5%. This further confirms the idea of stagnant growth in availability for low-resource language data.

All data was preprocessed by: (i) removing special characters such as equal signs, dollar signs and pound signs for the sake of clarity they are exemplified here "$ € = ;#", (ii) filtering out long sentences (more than 150 characters), and (iii) tokenizing sentences on spaces and punctuation using the Moses tokenizer tool [3].

[3] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/

Language	OpenSub2016	OpenSub2018	Increase (%)
Albanian (SQ)	3.0K	3.1K	+3.3%
Dutch (NL)	98.2K	125K	+27.3%
English (EN)	322K	447K	+38.8%

Table 1: Overview number of subtitle files for NL, EN and SQ in the OpenSubtitles 2016 and OpenSubtitles 2018 datasets, including the increase (%) of files.

Table 2 shows the tokens per language set. Table 2 shows the amount of parallel sentences obtained per language pair, and their corresponding number of tokens. In order to reduce the effect of differences in corpora sizes between high- and low-resource languages the Dutch-English pair was reduced to 2 million sentences from its original 37 million parallel sentences to match the English-Albanian corpora, as seen in Table 2. For the other language pairs (EN-SQ and NL-SQ), the maximum amount of data available was used. Additionally, the data was split 70/20/10 for training, development and testing. A batch of 100 NL-SQ sentences was sampled from the test set for human evaluation.

Language pair	Sentences	Tokens source	Tokens target
NL-SQ	1.6 M	12.4 M	13.2 M
EN-SQ	1.9 M	15.3 M	14.0 M
NL-EN	2.0 M	14.6 M	16.9 M

Table 2: Overview of the amount of parallel sentences and tokens available per language (pair).

3.2 Machine Translation Systems

Three Neural MT methods were trained and compared using the OpenNMT library (Klein et al., 2017): a one-to-one NL↔SQ model, a pivot translation model and a ZST/FST NMT model. For the implementation, we relied on the translation pipeline provided on GitHub by Shterionov (2018). For the Transformer systems we used OpenNMT-py.[4] The systems were trained for a maximum of 30K steps, saving an intermediate model every 1000 steps for 5 intermediate models. The options we used for the neural systems: number of layers *6*, size *256*, transformerff *2048*, number of heads *8*, dropout *0.1*, batch size *4096*, batch type *tokens*, learning optimizer *Adam* with $beta_2$=*0.998*, learning rate *2*. The Transformers have the learning rate decay enabled and the training data is distributed over a single Tesla P100-PCIE-16GB GPU powered by Google Colab. We use settings suggested by the OpenNMT community[5] as the optimal ones that lead to a quality on par with the original Transformer by Vaswani et al. (2017). Sub-word units (Sennrich et al., 2015) were used to build the vocabulary for the NMT systems, mitigating the out-of-vocabulary problem. We used BPE with 50k merging operation for all data sets.

The simplest model, i.e. the one-to-one NMT, is trained on the NL↔SQ data set. For the two-step pivot MT approach, two one-to-one models were trained: an NL↔EN model and an EN↔SQ model. The pivot approach requires two models for a one-way translation making it the least efficient approach. The final model ZST NMT is trained on the same data as the pivot approach but uses a single NMT model to translate between NL and SQ instead of two separate models as illustrated in Fig.1. Tokens indicating the translation direction per language (<2EN>, <2NL>, <2SQ>) are added to the training of the ZST model, allowing the specification of the

tokenizer.perl

[4] https://opennmt.net/OpenNMT-py/

[5] https://opennmt.net/OpenNMT-py/FAQ

desired target language at generation time.

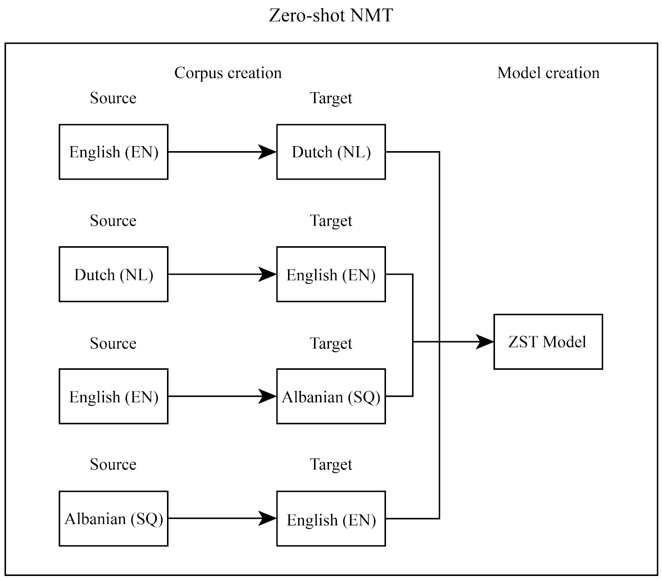

Figure 1: Zero-shot NL→EN NMT translation pipeline.

We additionally compare three ZST/FST models: ZST NMT, FST 50K NMT, and FST 150K NMT, where 50k and 150k of parallel NL→SQ data is added to the original ZST models, converting the system into a few-shot one. This way, we aim to measure the effect of adding the often limited available parallel.

3.3 Human evaluation

The human evaluation was conducted by three native Dutch and Albanian speakers that are also fluent in English. 100 sentences, varying in length, were sampled from the test sets of 1000 sentences. This evaluation serves as a supplement to the automatic evaluation methods, further analyzing bottlenecks missed by automatic evaluation metrics. The sentences were defined as either correct or incorrect. The correct section is divided into two categories: correct and correct without context. The correct without context section relates to generated translations being correct literal translations, or verbatim translations. These translations are highly dependent on the conversation topic, some translations can be considered accurate as they generate a sentence with the exact same wording as the source sentence. Yet, in some situations the generated translation may lack meaning as the context is not present in either the source, or the reference sentence. The incorrect section was divided into five error types: structural errors (word placement and sentence structure), missing words, incorrect word choice and incorrect language. The incorrect language category was introduced for analyzing the errors that occur during the transition of languages when using bridging methods.

4 Results

4.1 Automatic evaluation

As shown in Table 3, the highest performing model, according to the evaluation metrics, BLEU and TER, is the pivot model. Table 4 contains the results of the direct translations (NL→EN and EN→SQ). This table indicates how models perform on high-resource languages in comparison with low-resource languages. These values serve as a guideline to highlight how well the pivot MT performs on the basis of a low-resource language. As the pivot essentially combines the models in Table 4, into one model, it is beneficial to see how the performance is altered when running a pivot model from Dutch to Albanian, in contrast to the individual branches running directly.

Model	BLEU↑	TER↓
NL→SQ	13.68	0.65
pivot MT	**16.68**	**0.64**
ZST NMT	7.89	1.01
FST 50K NMT	10.98	0.96
FST 150K NMT	12.85	0.95

Table 3: Performance per model according to the automatic BLEU and TER metrics.

Model	BLEU↑	TER↓
NL→EN	33.41	0.46
EN→SQ	22.87	0.58

Table 4: BLEU and TER scores for the NL→EN & EN→SQ direct translations models.

As previously stated, this research compares three methods: one-to-one NMT, pivot MT, and ZST NMT. The pivot MT is the best performing method since both the NL→EN and EN→SQ (Table 3 and 4) achieve BLEU scores that indicate an acceptable translation quality.

While the ZST is the worst performing model, the addition of parallel data (50K and 150K) does increase its performance. However, this means that parallel data between the source and target language is necessary to produce acceptable results.

It is worth mentioning that while the pivot MT outperforms the one-to-one NL→SQ NMT by 3 points in terms of BLEU score, the TER score differs slightly. This matter could describe that while the NL→SQ produces less accurate sentences than the pivot MT, the number of edits required to transform the generated sentences into the reference sentences is close to equal.

4.2 Human evaluation

Table 5 gives an overview of the human evaluation in terms of correct/incorrect translations. Again, the pivot MT (NL→EN→SQ) appears to be the (overall) best performing system.

Model	Correct	Correct/Context	Total Correct
NL-SQ	62	8	70
pivot MT	63	13	76
ZST NMT	32	1	33
ZST 50K NMT	55	4	59
ZST 150K NMT	58	7	65

Table 5: Human Evaluation of 100 random sample sentences.

Table 6, shows the incorrect sentences from the human evaluation process. When it comes to spelling mistakes all models scored perfectly on this category and made no mistakes, this is due to the fact that the models base the spelling directly on the training data, meaning if there are no spelling mistakes in the dataset, the model will not make spelling mistakes on its

Models	Structure	Missing word	Word Choice	Language	Total Incorrect
NL-SQ	5	22	3	0	30
pivot MT	7	17	0	0	24
ZST NMT	5	34	9	19	67
FST 50K NMT	2	31	8	0	41
FST 150K NMT	4	28	3	0	35

Table 6: Overview of the detailed human evaluation, dividing the errors into different types: structure, missing words, word choice and language mistakes.

own. Incorrect language is only applicable to the ZST NMTs due to implicit bridging. In this situation the ZST model showed 19 incidents, within the selected 100 sentences, where words appeared in the wrong language, neither the source nor the target language, this issue is further discussed in the next section. This table also highlights the value of adding parallel corpora to the ZST model as it shows an overall improvement on the missing word error as well as the language choice error. Overall, the results strongly indicate to the pivot MT performing the best low-resource translation from Dutch to Albanian.

5 Discussion

According to the automatic and human evaluation, the pivot approach performs best, however, as evidenced in Table 4, adding parallel corpora to the ZST training data rapidly improves its performance. Low-resource languages often lack of training data and thus the ability to add parallel corpora may not always be present. Johnson et al. (2016), create a more promising analysis of a ZST model on low-resource languages, however, next to producing their own dataset, the amount of data available to Johnson et al. (2016) is far more substantial. Due to time constraints creating such a dataset and running it is out of the capabilities of this research. Additionally, Johnson et al. (2016) in contrast to this research, worked with single language pairs, operating with 255 million parameters per model, whereas this research operated on 55 million parameters per model, five times less the amount of Johnson et al. (2016).

A major point that requires addressing, is the difficulty of translating any language when the context is lacking. In this specific case the data used for the sentences came from movie translations, where context is inherently significant. Two examples that highlight the difficulty will be explored and discussed below.

NL Source	Het was verkeerd wat ze deden .
EN Translation	It was wrong what they did .
SQ Reference	Atë që kanë bër është gabim Valerie .
Pivot MT	Ishte gabim ajo që bënë ata.

Table 7: Translation generated by the pivot MT model for the Dutch sentence “Het was verkeerd wat ze deden.”

Table 7 illustrates an example of a translation produced by the pivot MT model. The Dutch input sentence “Het was verkeerd wat ze deden.” can be translated into English as “It was wrong what they did”, a translation that closely reflects the word placement whilst capturing the message of the phrase. The SQ reference sentence provided, can be literally translated into English as “What they did was wrong Valerie.”. The reference thus contains an additional word “Valerie” which is not present in the source. This could be due to the specific context in which this sentence was uttered. The translation generated by the pivot MT can be literally translated as "It was wrong what they did", a translation which not only follows the reference sentence verbatim but also carries forth the sentiment and message embedded in the sentence. This example illustrates some of the shortcomings of the automatic evaluation metrics while highlighting the importance of contextual cues and ambiguity in translation.

NL Source	Wat is oké ?
EN Translation	What is okay ?
SQ Reference	Çfarë është në rregull ? (EN: What is okay?)
One-To-One	Çfarë është ? (EN: What is?)
Pivot MT	Çfarë ke ? (EN: What's up?)
ZST MT	Çfarë është mirë ? (EN: What is okay?)

Table 8: Overview of translations generated by the one-to-one, pivot MT and ZST MT models for the Dutch sentence "What is oké"

This is, however, not the case for all sentences in a movie, as suggested by the examples in Table 8. This example highlights the importance of context in translation.

In Table 8, the ZST seems to generate the most accurate and fluent translation given the NL source sentence "What is oké?" (EN: "What is okay?"). The one-to-one model generates an incomplete translation while the pivot MT generated an incorrect translation. However, without any further context, and given the fact that the reference is rather vague, it is nearly impossibly to determine which one of those translations is the most accurate.

Finally, Table 6 presents the human evaluation of the models. In terms of the models this table reiterates the fact that the ZST model performed the worst in translating accurately. Furthermore, the ZST is the only model that made a language error, when translating from the reference to the generated sentence, some words came out in English rather than Albanian (see Appendix C: Sentence 13). This issue is explored in the Johnson et al. (2016) paper in relation to Japanese and Korean translation, by feeding a linear combination of the embedding vectors giving it a notation of 0 and 1. In the midst of the translation the model produces an output of 0.5, in some cases translating from Japanese to Korean, and in other instances with an output of 0.58 producing a mix of both languages resulting in an incoherent sentence, a situation that may be attributed to a difference in scripts. This investigation by Johnson et al. (2016) is relevant here as the multilingual ZST model used also resulted in some instances of mixed language outputs. In addition to the language error, the ZST model also performs the worst in terms of word choice in the generated translation, however, as posited in the second sub-question, adding parallel corpora improves the model accuracy. Overall, Table 6 restates the conclusion that the pivot-based NMT outperforms the other models when accurately translating Dutch to Albanian to the largest extent.

6 Conclusion

In this paper, three approaches to NL→SQ MT are explored: a one-to-one direct model and two approaches specific to low-resource settings, Pivot-NMT and ZST, including FST - where small amounts of parallel data was added to the ZST models. From our experiments it results that the pivot approach outperformed the others in terms of the automatic (BLEU & TER) and human assessment. Additional experiments were conducted where small amounts of parallel NL-SQ data was added to the ZST training data leading to improvements, approaching the results obtained using English as a pivot. Additionally, ZST/FST has some advantages over pivot-based MT in terms of efficiency as it only requires the training of one model. In future work, we would like to further explore how parallel data affects the performance of ZST models and experiment with different, morphologically richer pivot languages since English does not capture many of the specific linguistic properties of Albanian (gender, cases...).

References

Cheng, Y., Yang, Q., Liu, Y., Sun, M., and Xu, W. (2017). Joint training for pivot-based neural machine translation. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3974–3980.

Cieri, C., Maxwell, M., Strassel, S., and Tracey, J. (2016). Selection criteria for low resource language programs. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4543–4549.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2021). *Ethnologue: Languages of the World. Twenty-fourth edition*. Dallas, Texas: SIL International, Online version: http://www.ethnologue.com.

Ha, T.-L., Niehues, J., and Waibel, A. (2016). Toward multilingual neural machine translation with universal encoder and decoder. *Institute for Anthropomatics and Robotics*, 2(10.12):16.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., et al. (2016). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Kallulli, D. (2011). 9 albanian. In *The Languages and Linguistics of Europe*, pages 199–208. De Gruyter Mouton.

Klein, G., Kim, Y., Deng, Y., Senellart, J., and Rush, A. M. (2017). Opennmt: Open-source toolkit for neural machine translation. In *Proc. ACL*.

Koehn, P. and Knowles, R. (2017). Six challenges for neural machine translation. *ACL 2017*, page 28.

Lakew, S., Cettolo, M., and Federico, M. (2018). A comparison of transformer and recurrent neural networks on multilingual neural machine translation. In *27th International Conference on Computational Linguistics (COLING)*, pages 641–652.

Lison, P. and Tiedemann, J. (2016). Opensubtitles2016: Extracting large parallel corpora from movie and tv subtitles. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 923–929.

Lison, P., Tiedemann, J., and Kouylekov, M. (2018). Opensubtitles2018: Statistical rescoring of sentence alignments in large, noisy parallel corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Liu, C.-H., Silva, C. C., Wang, L., and Way, A. (2018). Pivot machine translation using chinese as pivot language. In *China Workshop on Machine Translation*, pages 74–85. Springer.

Magueresse, A., Carles, V., and Heetderks, E. (2020). Low-resource languages: A review of past work and future challenges. *arXiv preprint arXiv:2006.07264*.

Mattoni, G., Nagle, P., Collantes, C., and Shterionov, D. (2017). Zero-shot translation for indian languages with sparse data. *Proceedings of the 16th machine translation summit (MTSummit 2017)*, 2:1–10.

Prifti, P. R. (2008). Albanian literature. *Translation Review*, 76(1):29–31.

Psacharopoulos, G. and Woodhall, M. (1993). *Education for development*. Citeseer.

Pustina, B. (2016). Transmitting albanian cultural identity in the age of the internet. *New Review of Information Networking*, 21(1):24–36.

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Shterionov, D. (2018). Nmtscripts. `https://github.com/dimitarsh1/NMTScripts`.

Singh, A. K. (2008). Natural language processing for less privileged languages: Where do we come from? where are we going? In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*.

Tapo, A. A., Coulibaly, B., Diarra, S., Homan, C., Kreutzer, J., Luger, S., Nagashima, A., Zampieri, M., and Leventhal, M. (2020). Neural machine translation for extremely low-resource african languages: A case study on bambara. In *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, pages 23–32.

Teplitskiy, M., Lu, G., and Duede, E. (2017). Amplifying the impact of open access: Wikipedia and the diffusion of science. *Journal of the Association for Information Science and Technology*, 68(9):2116–2127.

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 *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010.