

# Task-Oriented Dialog Systems for the Senegalese Wolof Language

Derguene Mbaye<sup>1,2</sup>

<sup>1</sup>Baamtu Technologies

<sup>2</sup>Université Cheikh Anta Diop

Dakar, Sénégal

derguenembaye@esp.sn

Moussa Diallo<sup>3</sup>

<sup>3</sup>Ecole Supérieure Polytechnique (ESP)

Dakar, Sénégal

moussa.diallo@esp.sn

## Abstract

In recent years, we are seeing considerable interest in conversational agents with the rise of large language models (LLMs). Although they offer considerable advantages, LLMs also present significant risks, such as hallucination, which hinder their widespread deployment in industry. Moreover, low-resource languages such as African ones are still underrepresented in these systems limiting their performance in these languages. In this paper, we illustrate a more classical approach based on modular architectures of Task-oriented Dialog Systems (ToDS) offering better control over outputs. We propose a chatbot generation engine based on the Rasa framework and a robust methodology for projecting annotations onto the Wolof language using an in-house machine translation system. After evaluating a generated chatbot trained on the Amazon Massive dataset, our Wolof Intent Classifier performs similarly to the one obtained for French, which is a resource-rich language. We also show that this approach is extensible to other low-resource languages, thanks to the intent classifier's language-agnostic pipeline, simplifying the design of chatbots in these languages.

## 1 Introduction

Artificial Intelligence (AI) has experienced a tremendous growth in recent years, mainly due to the rapid development of Deep Learning. The latter has allowed to achieve super-human skills on tasks such as Image Classification. This has been mainly made possible thanks to the development of large datasets and the considerable increase in computing power and their accessibility. NLP has thus leveraged these advances and has had its "ImageNet moment"<sup>1</sup> with the emergence of huge corpora as well as the possibility of leveraging pre-trained models to apply them to downstream tasks through Transfer Learning (Ruder et al., 2019).

However, the availability of such corpora only concerns a small group of languages such as English, Chinese or French which are referred to as "Resource Rich". The majority of the approximately remaining 7,000 languages (Eberhard et al., 2019), and particularly African ones, fall into the "Low Resource" category and struggle to have sufficient corpora and NLP tools (Hedderich et al., 2021). They are thus left behind in most of the AI revolutions, such as LLMs, which reach the state of the art in many NLP tasks but struggle to reproduce equivalent performance in African languages. A human-translated benchmark dataset for 16 typologically diverse low-resource African languages (including Wolof) has been presented by (Adelani et al., 2024) covering three tasks: natural language inference, mathematical reasoning, and multi-choice knowledge-based QA. They evaluated zero-shot, few-shot, and translate-test settings (where test sets are translated into English) across 10 open and four proprietary LLMs and revealed a significant performance gap between high-resource languages (English and French) and African ones. A lot of work has therefore been done to address this challenge and a great illustration is the emergence of initiatives such as the Masakhane<sup>2</sup> community which brings together thousands of researchers, practitioners, linguists and enthusiasts to produce datasets and models for a wide range of African languages (Nekoto et al., 2020). Other data collection projects have also included African languages, and especially Wolof, such as (Tiedemann, 2012; Strassel and Tracey, 2016; Goyal et al., 2022; Federmann et al., 2022).

However, the data collection process is a very time-consuming task and can quickly become tedious. Additionally, for some languages, the text may not even exist in electronic form, or worse still, the text may be obtained from phonetic tran-

<sup>1</sup><https://www.ruder.io/nlp-imagenet/>

<sup>2</sup><https://www.masakhane.io/>

criptions directly from audio, in the absence of a defined written standard. So, it is important to point out here that not all low-resource languages are at the same level in terms of resource scarcity. Other initiatives have therefore studied the usability of work already done in resource-rich languages in order to apply it to low resource ones. This is particularly the case with Cross-Lingual Transfer which is an approach where knowledge gained while training a model on a particular language (or set of languages) is leveraged to improve the performance on tasks in a different one (Artetxe et al., 2020; Lample and Conneau, 2019). This state-of-the-art technique is applicable to virtually all natural language processing (NLP) tasks, including Task-oriented Dialog Systems (ToDS) where the lack of datasets in most languages for both training and evaluation is the most critical factor preventing the creation of truly multilingual ToDS (Razumovskaia et al., 2022). ToDS are software programs (or agents) that users talk to in order to carry out a given task. To do this, they usually start by identifying the user’s goal, often called intent, and the associated arguments, called slots, before being able to process the request appropriately. The most common applications are hotel, taxi- and restaurant reservations, information supply and customer support, among others. In the following example: **Book me a room for a person from July 15 to July 24**, the intent would be `book_room` and the `start_date` and `end_date` slots would have the values `July 15` and `July 24` respectively. Intent recognition and slot filling are thus the main tasks involved in a ToDS, and can be done either separately (Arora et al., 2020) or jointly (Schuster et al., 2019) in a single task. This architecture, in which the system is divided into several modules performing specific tasks, is called modular architecture as opposed to the end-to-end one which uses a neural network for all these tasks (Zhang et al., 2020). This makes them very data-intensive, as they need to model several different tasks from training data.

In this paper, we propose to leverage the benefits of Cross Lingual Transfer to build a Wolof Chatbot generation engine based on an in-house French↔Wolof machine translation system and a language agnostic pipeline for intent classification and slot filling using Rasa (Bocklisch et al., 2017). We introduce a simple but effective approach to projecting annotations from a source language to a target one, using only the original machine translation

system coupled with an intuitive parsing strategy. The paper is therefore structured as follows:

- We begin by presenting some work done in Cross-lingual Transfer for Task oriented Dialog Systems on low-resource languages in Section 2.
- The Wolof language and the datasets used are presented in Section 3.
- Our annotation projection approach is presented in section 4.
- In Section 5, we present the different experiment settings.
- Section 6 presents the results.
- Conclusion and perspectives are presented in section 7.

## 2 Related Work

ToDS for low-resource languages has been the subject of extensive research in the community and various approaches have been studied. A first experience of a Wolof chatbot was proposed in (Gauthier et al., 2022) with a POC of a voice assistant based on Rasa and designed with manually collected synthetic data. This limits the ability to collect large volumes of data, which is the reason why more efficient alternatives have been explored.

### 2.1 Word alignment approach

Training data translation (translate-train) into target languages is an approach that is increasingly being considered to enhance the performance of cross-lingual transfer. (Schuster et al., 2019) explored three different cross-lingual transfer methods: translation of training data with a bidirectional neural machine translation system (NMT) following the work carried out by (McCann et al., 2018), using cross-lingual pre-trained embeddings, and a method of using a multilingual machine translation encoder as contextual word representations. Although the latter two showed better performance in the presence of hundreds of training data, in cases where no data is available in the target language, translating the training data gives the best results. This translation approach has also been explored by (López de Lacalle et al., 2020), which leveraged the Spanish subset of the dataset of (Schuster et al., 2019) to translate it into Basque using a Spanish-to-Basque Transformer-based NMT system (Vaswani

et al., 2023) before projecting the annotations using word alignment (Dyer et al., 2013). Two additional models were used to perform the word alignment, making this approach more laborious. A similar approach was also presented by (Kanakagiri and Radhakrishnan, 2021) with an NMT-based dataset creation approach and an annotation projection based on token prefix matching and mBERT (Devlin et al., 2019) based semantic matching from a rich resource language. For the translation phase, they also experimented with a transformer-based model but also with Google Translate services.

## 2.2 Marker-based approach

To avoid the need for additional models for word alignment, other simpler alternatives are considered, such as mark-then-translate as explored in (Lewis et al., 2020). This approach consists of inserting specific markers around the corresponding chunks in the source sentence (e.g., [marker] and [/marker]), then translating into the target language to maintain these markers during the translation process. An optimized version of the mark-then-translate approach, has been proposed in (Chen et al., 2023) consistently outperforming the alignment-based approach. Authors used language-agnostic square bracket markers, combined with an efficient fine-tuning strategy of the NLLB (No Language Left Behind) model (Team et al., 2022) to encourage the multilingual machine translation system to better preserve the special markers during translation. They also showed that the alignment-based methods are more error-prone when projecting span-level annotations, compared to the marker-based approaches. However, inserting markers into the source sentence tends to compromise the translation quality (Chen et al., 2023). A different approach called Translate-and-Fill (TaF) is explored in (Nicosia et al., 2021), as opposed to Translate-Align-Project (TAP), which uses alignment and projection modules. TaF consists of a sequence-to-sequence filler model that constructs a full parse conditioned on an utterance and a view of the same parse. This approach requires however two seq2seq models trained differently: one is the usual semantic parser and the other is what they call the filler. Researchers in (Xu et al., 2020) introduced a single end-to-end model that learns to align and predict target slot labels jointly for cross-lingual transfer.

## 3 Data

### 3.1 The Wolof Language

Belonging to the Atlantic group of the Niger-Congo language family, Wolof is an African language mainly spoken in Senegal (the majority of the population) but also in Gambia and some parts of Mauritania. It is a non-tonal agglutinative language whose alphabet is quite close to the French one: we can find all the letters of its alphabet except H, V and Z (Mbaye et al., 2024). As with many African languages, Wolof is a low-resource language, as opposed to high-resource languages like English and Chinese. Although there is no consensus on a suitable definition of the low-resource concept, several researchers have explored definitions from different angles even beyond the lack of data and NLP tools. Thus, this concept has been defined in (Besacier et al., 2014) as a language that lacks a unique writing system, has a low Internet presence, and lacks linguistic expertise, among other things. Wolof linguistics has, however, benefited from a lot of research but the lack of a unique writing system makes the design of NLP tools particularly challenging in this language. There is a writing system based on the Latin script and another one based on the Arabic script called Wolofal or Ajami (Mbaye et al., 2024). The latter is however very little used and is generally located within usage related to the Islamic religion which is the predominant religion in Senegal. The Latin script, on the other hand, is the most widespread, and is the one we'll be considering in this work. However, due to a lack of language standardization, two forms of writing illustrated in (Mbaye and Diallo, 2023) are noted with the Latin script:

1. A so-called official form, respecting the published alphabet and adopted by all Wolof datasets open to date;
2. A conventional form that doesn't respect any rules and is very present on media such as social networks.

We'll therefore consider the official form in this article.

### 3.2 Dataset

There are several multilingual datasets in Task-oriented dialog systems, typically for intent classification and slot filling. Some datasets specialize in a particular domain, such as MultiATIS++ (Xu et al.,

2020), which is specific to air travel, while others are multi-domain, such as MTOP (Li et al., 2021), which covers 11 domains. Among ToDS datasets, the MASSIVE dataset (FitzGerald et al., 2023) offers the greatest diversity in terms of languages and domains, with 51 languages and 18 domains covered, as illustrated in the Slot and Intent Detection benchmark and datasets presented in (Kwon et al., 2023). It also contains examples in French, which will enable us to use a French→Wolof machine translation system to project annotations in Wolof. Containing 1M labelled realistic utterances and around 16,500 utterances in French, we extracted 10 domains out of 18 spread over 27 intents and around 10,000 utterances (examples), as illustrated in Table 1. French is the easiest language to work with when it comes to leveraging a resource-rich language to perform NLP tasks on Wolof. Since the official language in Senegal is French, it will be easier to find translators and other linguists on the French-Wolof language pair, and most existing translation datasets are in this language pair too. Wolof spoken in Senegal is also very much code-switched with French, with many French words used in everyday conversations.

#### 4 Annotation Projection

Annotation projection involves transferring the labels present in a sentence in a source language to the translated sentence in the target language. A machine translation system is thus at the heart of the process, and the resource-limited nature of the Wolof language adds a further layer of complexity. A French Wolof machine translation system based on LSTMs was presented in (Mbaye et al., 2024), and another one based on Vanilla Transformers in (Dione et al., 2022). But pre-trained multilingual translation models have shown the most interesting performances in supporting low-resource languages. They allow information sharing between similar languages that greatly allows to improve translation on the language pairs as studied in (Ari-vazhagan et al., 2019). A lot of work has been done in this direction and a wide range of multilingual translation models have therefore been developed. Researchers in (Adelani et al., 2022) compared a set of multilingual models for machine translation in over twenty African languages, including Wolof, and the M2M100 (Fan et al., 2020) model showed the best performance. A distilled version of this model has been proposed in (Mo-

Domains	Intents	No examples
transport	transport_query	314
	transport_ticket	187
	transport_taxi	150
	transport_traffic	154
calendar	calendar_query	794
	calendar_set	1150
	calendar_remove	426
alarm	alarm_set	254
	alarm_remove	113
	alarm_query	183
lists	lists_query	299
	lists_remove	253
	lists_createoradd	241
takeaway	takeaway_query	181
	takeaway_order	177
play	play_audiobook	226
	play_game	169
	play_music	938
	play_podcasts	290
	play_radio	401
news	news_query	709
recommendation	recommendation_locations	235
	recommendation_events	259
	recommendation_movies	102
datetime	datetime_query	502
	datetime_convert	76
weather	weather_query	855
<b>Total No examples</b>		<b>9638</b>

Table 1: Description of the Massive dataset extract selected

hammadshahi et al., 2022), offering equivalent performance while being 3.6x smaller and 4.3x faster at inference. We finetuned this model simultaneously in both directions on an in-house dataset of 175,000 French-Wolof sentences and obtained a BLEU score of **26.38**.

We then empirically tested a set of markers including xml tags, dollars, braces, brackets, parentheses and some random special characters on sample translations, but found a range of issues:

1. These markers tend to change the meaning of the sentence ;
2. Their consistency is poor; some markers are preserved during translation in some sentences, but not in others, with no apparent pattern ;
3. When they are preserved, the labels surrounded by the markers are sometimes also translated, which we aim to avoid.



Figure 1: The three-step annotation projection algorithm: Parsing, which replaces the source annotations with ids, translation of the parsed sentence and backfilling of the translated annotations.

To address these challenges, we opted for “identifiers” instead of markers, by completely replacing the elements to be projected with numbered labels. Some numbers used as identifiers ended up being converted to letters during translation, reproducing the consistency issue mentioned above. We therefore combined the two approaches (markers and identifiers) by testing the markers presented earlier on the numbered identifiers. In the end, we found that only the dollar sign (\$) managed to preserve the marked identifiers during translation on all the sample sentences we tested. This may be due to the fact that this combination is the only one that is rare enough in the training data for the model not to confuse it and preserve it instead of trying to translate it in a certain way. We thus replaced all the labeled words with an identifier of the form \$0N\$, where N is a number ranging from 0 to the total number of labeled terms in the dataset. The label of each replaced word is stored in a dictionary, as is the translation of the word in question. The resulting sentences were then passed to the translation system to obtain the equivalents in the target language (Wolof) while keeping the identifiers intact during the translation process. From there, the identifiers in the translated sentences are replaced back by the translation of the words stored in the dictionary with their labels. Thus, we obtain as output the translated sentences with their corresponding annotations in the target language. The overall procedure is illustrated in Fig. 1.

## 5 Experiments

Instead of directly building an intent classifier and slot filling, we propose a system for generating chatbots on the fly (chatbot engine), based on the Rasa framework (Bocklisch et al., 2017). Rasa is an open-source framework designed for creating conversational AI chatbots. It offers tools and libraries for building and deploying AI-powered, text and

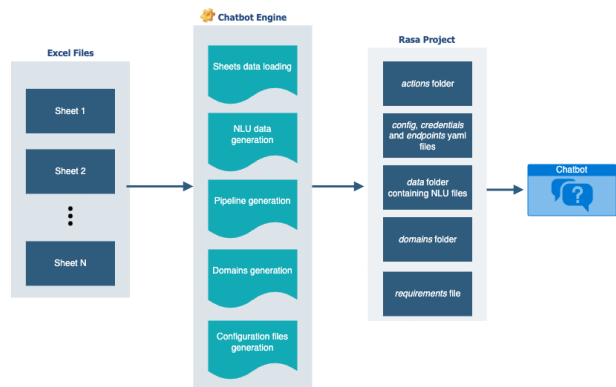


Figure 2: Diagram of the chatbot engine’s processing of excel files to output Rasa projects. Each Excel file constitutes a domain containing several sheets corresponding to intents, and each sheet contains the intent’s example data.

voice based chatbots capable of engaging in natural language conversations with users. The chatbot engine takes as input one or more Excel files which form the bot’s ontology (domain, nlu and dialog data), and then generate a working RASA Chatbot from this as illustrated in Fig. 2.

A benchmark for evaluating language-agnostic intent classification has been studied in (Wang et al., 2022) and Language-agnostic BERT Sentence Embedding (LaBSE) (Feng et al., 2022) produced the highest accuracy in almost all evaluation settings. Despite the long training times involved, this model allows the chatbots designed on top of it to be particularly flexible in terms of language support. We’ve used it to propose a simple but highly effective fixed pipeline, as illustrated in Fig. 3.

This pipeline is generated by the chatbot engine regardless of the language used in the excel files, making the overall system scalable to other languages beyond Wolof. At the heart of the pipeline is DIET<sup>3</sup> (Bunk et al., 2020), a state-of-the-art Rasa model for intent detection and slot filling. Since French is a resource-rich language, it is better represented in LaBSE than Wolof, enabling it to create richer embeddings and thus a more efficient intent classifier. We therefore use the French part of the dataset as a baseline to evaluate performance on synthetic data generated through translation and annotation projection. We then randomly split our dataset into train/test with a ratio of 80/20. The results are reported as F1 scores.

<sup>3</sup>Dual Intent and Entity Transformer

```

pipeline:
- name: WhitespaceTokenizer
- name: LanguageModelFeaturizer
  model_name: bert
  model_weights: "rasa/LaBSE"
- name: RegexFeaturizer
- name: DIETClassifier
  epochs: 100
  learning_rate: 0.001
- name: EntitySynonymMapper
- name: ResponseSelector
  epochs: 100

```

Figure 3: Pipeline of user input processing modules defined in the config.yml file generated by the chatbot engine.

## 6 Results

The results of the intent classification are shown in Table 2. We show the macro F1 scores on the source and synthetic data and we can observe that we get equivalent scores on both sides. This shows that the model succeeds in discriminating the intents in the synthetic dataset, suggesting a sufficiently qualitative translation. However, the model seems less confident when it comes to Wolof predictions, as shown in Fig.4 and Fig.5.

Confidence scores for good predictions vary more in the synthetic dataset than in the source dataset. This shows that, after translation, the model is more likely to confuse certain intentions, as illustrated by Fig.6. The calendar\_query and reccomandation\_events intents, for example, show confusion even though they are quite distinct. This can be attributed to the translation system, which may have had difficulty in producing accurate translations in some cases.

Table 3 illustrates the performance of annotation projection on the two datasets French and Wolof expressed in micro and macro F1 score as well as in accuracy. We note a pronounced discrepancy between the two datasets, with predictions on the synthetic data lower than those on the source data. It is important to point out that some of the annotations in the French dataset are expressions (not words) and are translated separately from the base sentence and therefore from the original context.

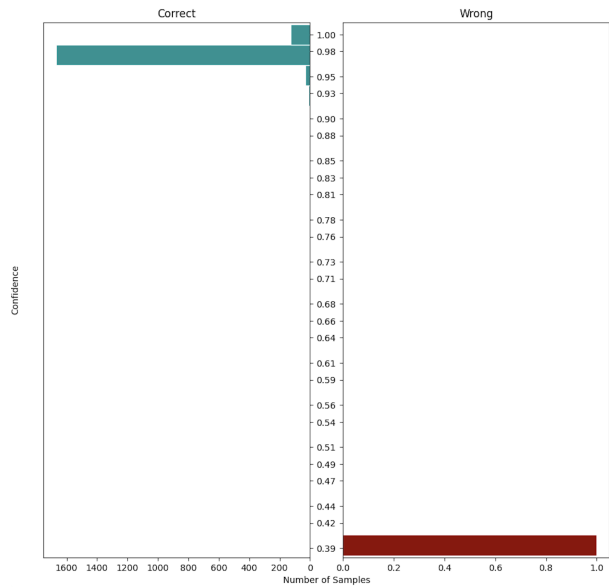


Figure 4: Intent prediction confidence distribution on the French dataset

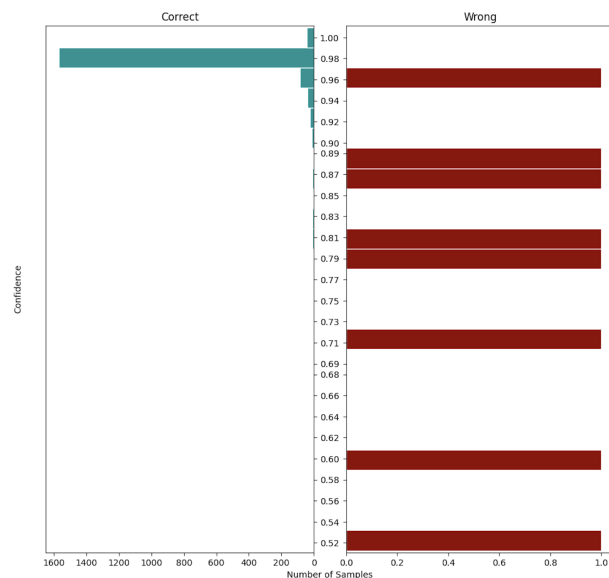


Figure 5: Intent prediction confidence distribution on the Wolof dataset

Intents	French	Wolof
transport_query	1.0	1.0
transport_ticket	1.0	1.0
transport_taxi	1.0	1.0
transport_traffic	1.0	0.98
calendar_query	0.99	0.99
calendar_set	1.0	0.99
calendar_remove	0.99	1.0
alarm_set	1.0	0.98
alarm_remove	1.0	1.0
alarm_query	1.0	1.0
lists_query	1.0	0.99
lists_remove	1.0	0.97
lists_createoradd	1.0	0.98
takeaway_query	1.0	1.0
takeaway_order	1.0	1.0
play_audiobook	1.0	1.0
play_game	1.0	1.0
play_music	1.0	1.0
play_podcasts	1.0	1.0
play_radio	1.0	1.0
news_query	1.0	0.99
recommendation_locations	1.0	1.0
recommendation_events	1.0	0.97
recommendation_movies	1.0	1.0
datetime_query	1.0	0.99
datetime_convert	1.0	1.0
weather_query	1.0	1.0
<b>macro avg</b>	<b>0.999</b>	<b>0.995</b>

Table 2: Report of the F1 scores of the intent classification on the French extract from the Massive dataset and its translation into Wolof

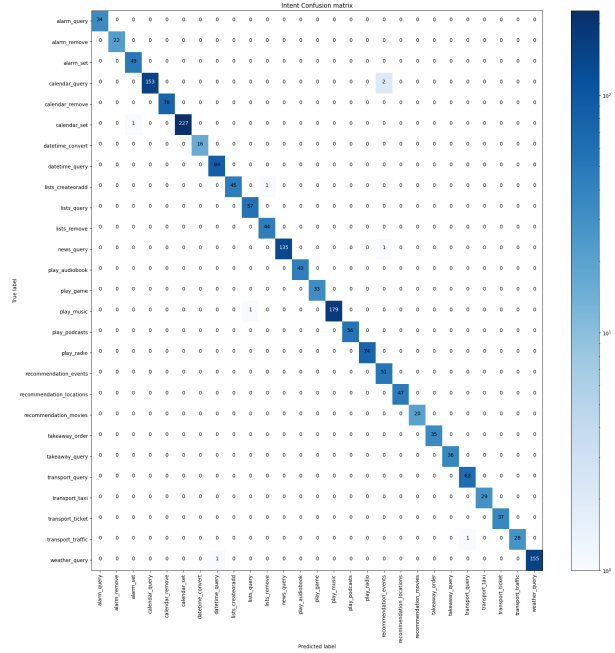


Figure 6: Intent confidence matrix on the Wolof dataset

	French	Wolof
<b>micro avg</b>	0.97	0.89
<b>macro avg</b>	0.96	0.86
<b>accuracy</b>	0.98	0.94

Table 3: Slot filling performance on French and Wolof datasets in micro and macro F1 Score and accuracy

This can affect the quality of the translation and the ability of the final model to discriminate between different labels. Generally speaking, annotation projection is heavily influenced by the machine translation system, which needs to be optimized as much as possible to create high-quality synthetic datasets.

## 7 Conclusion and Perspectives

In this paper, we illustrated an efficient approach to building dialogue systems in a low-resource language. We have shown how to leverage machine translation systems to create synthetic datasets using our annotation projection method. Our experiments showed that training on these synthetic datasets in Wolof gave competitive results compared with the French source data, which is a resource-rich language. However, this approach is strongly affected by the quality of the translation system, and translating annotations out of their original contexts can reduce the final quality of the

dataset. In the future, we will study techniques for improving the quality of the projection, as well as improving the representativeness of Wolof in a language model such as LaBSE. We will also study data augmentation approaches, which are particularly important when addressing low-resource domains. We will also study the integration of spelling correction systems to take into account the variations in writing commonly observed in a language like Wolof.

## Acknowledgment

This work is supported by the Google PhD Fellowship program.

## References

- David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruitter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajudeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencía Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. [A few thousand translations go a long way! leveraging pre-trained models for African news translation](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3053–3070, Seattle, United States. Association for Computational Linguistics.
- David Ifeoluwa Adelani, Jessica Ojo, Israel Abebe Azime, Jian Yun Zhuang, Jesujoba O. Alabi, Xuanli He, Millicent Ochieng, Sara Hooker, Andiswa Bukula, En-Shiun Annie Lee, Chiamaka Chukwuneke, Happy Buzaaba, Blessing Sibanda, Godson Kalipe, Jonathan Mukiibi, Salomon Kabongo, Foutse Yuehgoh, Mmasibidi Setaka, Lolwethu Ndolela, Nkiruka Odu, Rooweither Mabuya, Shamsuddeen Hassan Muhammad, Salomey Osei, Sokhar Samb, Tadesse Kebede Guge, and Pontus Stenertorp. 2024. [Irokobench: A new benchmark for african languages in the age of large language models](#). *Preprint*, arXiv:2406.03368.
- Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George Foster, Colin Cherry, Wolfgang Macherey, Zhifeng Chen, and Yonghui Wu. 2019. [Massively multilingual neural machine translation in the wild: Findings and challenges](#). *Preprint*, arXiv:1907.05019.
- Abhinav Arora, Akshat Shrivastava, Mrinal Mohit, Lorena Sainz-Maza Lecanda, and Ahmed Aly. 2020. [Cross-lingual transfer learning for intent detection of covid-19 utterances](#).
- Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2020. [On the cross-lingual transferability of monolingual representations](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4623–4637, Online. Association for Computational Linguistics.
- Laurent Besacier, Etienne Barnard, Alexey Karpov, and Tanja Schultz. 2014. [Automatic speech recognition for under-resourced languages: A survey](#). *Speech Communication*, 56:85–100.
- Tom Bocklisch, Joey Faulkner, Nick Pawlowski, and Alan Nichol. 2017. [Rasa: Open source language understanding and dialogue management](#). *Preprint*, arXiv:1712.05181.
- Tanja Bunk, Daksh Varshneya, Vladimir Vlasov, and Alan Nichol. 2020. [Diet: Lightweight language understanding for dialogue systems](#). *Preprint*, arXiv:2004.09936.
- Yang Chen, Chao Jiang, Alan Ritter, and Wei Xu. 2023. [Frustratingly easy label projection for cross-lingual transfer](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5775–5796, Toronto, Canada. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Cheikh M. Bamba Dione, Alla Lo, Elhadji Mamadou Nguer, and Sileye Ba. 2022. [Low-resource neural machine translation: Benchmarking state-of-the-art transformer for Wolof->French](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6654–6661, Marseille, France. European Language Resources Association.
- Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. [A simple, fast, and effective reparameterization of IBM model 2](#). In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.
- David Eberhard, Gary Simons, and Chuck Fennig. 2019. *Ethnologue: Languages of the World, 22nd Edition*. SIL International.



- Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. 2020. [Beyond english-centric multilingual machine translation](#). *Preprint*, arXiv:2010.11125.
- Christian Federmann, Tom Kocmi, and Ying Xin. 2022. [NTREX-128 – news test references for MT evaluation of 128 languages](#). In *Proceedings of the First Workshop on Scaling Up Multilingual Evaluation*, pages 21–24, Online. Association for Computational Linguistics.
- Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. [Language-agnostic bert sentence embedding](#). *Preprint*, arXiv:2007.01852.
- Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, Swetha Ranganath, Laurie Crist, Misha Britan, Wouter Leeuwis, Gokhan Tur, and Prem Nataraajan. 2023. [MASSIVE: A 1M-example multilingual natural language understanding dataset with 51 typologically-diverse languages](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4277–4302, Toronto, Canada. Association for Computational Linguistics.
- Elodie Gauthier, Papa Séga Wade, Thierry Moudenc, Patrice Collen, Emilie De Neef, Oumar Ba, Ndeye Khoyane Cama, Ahmadou Bamba Kebe, Ndeye Aissatou Gningue, and Thomas Mendo’O Aristide. 2022. [Preuve de concept d’un bot vocal dialoguant en wolof \(proof-of-concept of a voicebot speaking Wolof\)](#). In *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, pages 403–412, Avignon, France. ATALA.
- Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc’Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. [The Flores-101 evaluation benchmark for low-resource and multilingual machine translation](#). *Transactions of the Association for Computational Linguistics*, 10:522–538.
- Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. 2021. [A survey on recent approaches for natural language processing in low-resource scenarios](#). *Preprint*, arXiv:2010.12309.
- Tushar Kanakagiri and Karthik Radhakrishnan. 2021. [Task-oriented dialog systems for Dravidian languages](#). In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 85–93, Kyiv. Association for Computational Linguistics.
- Sang Yun Kwon, Gagan Bhatia, Elmoatez Billah Nagoudi, Alcides Alcoba Inciarte, and Muhammad Abdul-mageed. 2023. [SIDLR: Slot and intent detection models for low-resource language varieties](#). In *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, pages 241–250, Dubrovnik, Croatia. Association for Computational Linguistics.
- Guillaume Lample and Alexis Conneau. 2019. [Cross-lingual language model pretraining](#). *Preprint*, arXiv:1901.07291.
- Patrick Lewis, Barlas Oguz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2020. [MLQA: Evaluating cross-lingual extractive question answering](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7315–7330, Online. Association for Computational Linguistics.
- Haoran Li, Abhinav Arora, Shuohui Chen, Anchit Gupta, Sonal Gupta, and Yashar Mehdad. 2021. [MTOP: A comprehensive multilingual task-oriented semantic parsing benchmark](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2950–2962, Online. Association for Computational Linguistics.
- Maddalen López de Lacalle, Xabier Saralegi, and Iñaki San Vicente. 2020. [Building a task-oriented dialog system for languages with no training data: the case for Basque](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2796–2802, Marseille, France. European Language Resources Association.
- Derguene Mbaye and Moussa Diallo. 2023. [Beqi: Revitalize the senegalese wolof language with a robust spelling corrector](#). *Preprint*, arXiv:2305.08518.
- Derguene Mbaye, Moussa Diallo, and Thierno Ibrahima Diop. 2024. [Low-resourced machine translation for senegalese wolof language](#). In *Proceedings of Eighth International Congress on Information and Communication Technology*, pages 243–255, Singapore. Springer Nature Singapore.
- Bryan McCann, James Bradbury, Caiming Xiong, and Richard Socher. 2018. [Learned in translation: Contextualized word vectors](#). *Preprint*, arXiv:1708.00107.
- Alireza Mohammadshahi, Vassilina Nikoulina, Alexandre Berard, Caroline Brun, James Henderson, and Laurent Besacier. 2022. [Small-100: Introducing shallow multilingual machine translation model for low-resource languages](#). *Preprint*, arXiv:2210.11621.
- Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Tajudeen Kolawole, Taiwo Fagbohunge, Solomon Oluwole Akinola, Shamsuddeen Hassan Muhammad, Salomon Kabongo, Salomey Osei, Sackey Freshia, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene

- Ahia, Musie Meressa, Mofe Adeyemi, Masabata Mokgesi-Seling, Lawrence Okegbemi, Laura Jane Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreuzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkabir Dangana, Herman Kamper, Hady Elshar, Goodness Duru, Gholah Kioko, Espoir Murhabazi, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Emezue, Bonaventure Dossou, Blessing Sibanda, Blessing Itoro Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. [Participatory research for low-resourced machine translation: A case study in african languages](#). *Preprint*, arXiv:2010.02353.
- Massimo Nicosia, Zhongdi Qu, and Yasemin Altun. 2021. [Translate & Fill: Improving zero-shot multilingual semantic parsing with synthetic data](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3272–3284, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Evgeniia Razumovskaia, Goran Glavaš, Olga Majewska, Edoardo M. Ponti, Anna Korhonen, and Ivan Vulić. 2022. [Crossing the conversational chasm: A primer on natural language processing for multilingual task-oriented dialogue systems](#). *Preprint*, arXiv:2104.08570.
- Sebastian Ruder, Matthew E. Peters, Swabha Swayamdipta, and Thomas Wolf. 2019. [Transfer learning in natural language processing](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*, pages 15–18, Minneapolis, Minnesota. Association for Computational Linguistics.
- Sebastian Schuster, Sonal Gupta, Rushin Shah, and Mike Lewis. 2019. [Cross-lingual transfer learning for multilingual task oriented dialog](#). 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)*, pages 3795–3805, Minneapolis, Minnesota. Association for Computational Linguistics.
- Stephanie Strassel and Jennifer Tracey. 2016. [LORELEI language packs: Data, tools, and resources for technology development in low resource languages](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3273–3280, Portorož, Slovenia. European Language Resources Association (ELRA).
- NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Hefernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. [No language left behind: Scaling human-centered machine translation](#). *Preprint*, arXiv:2207.04672.
- Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey. European Language Resources Association (ELRA).
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. [Attention is all you need](#). *Preprint*, arXiv:1706.03762.
- Gengyu Wang, Cheng Qian, Lin Pan, Haode Qi, Ladislav Kunc, and Saloni Potdar. 2022. [Benchmarking language-agnostic intent classification for virtual assistant platforms](#). In *Proceedings of the Workshop on Multilingual Information Access (MIA)*, pages 69–76, Seattle, USA. Association for Computational Linguistics.
- Weijia Xu, Batool Haider, and Saab Mansour. 2020. [End-to-end slot alignment and recognition for cross-lingual NLU](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5052–5063, Online. Association for Computational Linguistics.
- Zheng Zhang, Ryuichi Takanobu, Qi Zhu, Minlie Huang, and Xiaoyan Zhu. 2020. [Recent advances and challenges in task-oriented dialog system](#). *Preprint*, arXiv:2003.07490.