Low-resource Machine Translation for Sudanese Dialect

Mohamed Elabbas
African Institute for Mathematical Sciences
melabbas@aimsammi.org

Mohammed Abukalam
African Institute for Mathematical Sciences
mabukalam@aismammi.org

Louis Martin
Meta AI Research
louismartin@meta.com

Abstract

Addressing the prevalent challenge of language resource disparity, this research primarily aims to enhance the accessibility of English content for speakers of the Sudanese dialect, a recognized low-resource language. Our strategic approach involved fine-tuning an established English-Arabic multilingual machine translation model, thereby leveraging the existing linguistic similarities between the Sudanese dialect and Modern Standard Arabic. To achieve this, a distinct dataset was developed, comprised of 911 Flores-101 sentences translated into the Sudanese dialect. This dataset was instrumental in fine-tuning the model and in subsequent performance evaluation. The study was predicated on the M2M100 model, a highly proficient base model selected for its potential to significantly improve in handling the Sudanese dialect. A comparative analysis was conducted between the enhanced M2M100 model's output and that of advanced language models, namely Chat GPT, to evaluate performance with respect to low-resource languages. In an additional bid to assess the model's robustness and versatility, its effectiveness was evaluated across resource-rich language pairs, specifically English-Arabic, in bi-directional translation tasks.

1 Introduction

Neural Machine Translation (NMT) systems have made remarkable advancements in recent years, achieving state-of-the-art performance in numerous language pairs. This progress has piqued the interest of researchers who now question whether these current systems can rival the proficiency of human translators. While modern MT systems are typically trained on vast bilingual datasets, the availability of such data is not uniform across all language pairs. Many languages worldwide, including the Sudanese dialect, face the challenge of limited or non-existent resources for training translation models. These languages are categorized as low-resource languages, characterized by a scarcity of available training data.

Translation plays a vital role in enabling information accessibility, communication, and the preservation of culture and heritage in low-resource languages. It serves as a catalyst for

Preprint. Under review.

promoting education and literacy within these communities, leading to their development and empowerment. Furthermore, translation acts as a bridge, connecting diverse communities by overcoming language and cultural barriers, fostering understanding, and nurturing tolerance. In the specific context of the Sudanese dialect, translation endeavors are instrumental in preserving its unique linguistic and cultural identity, ensuring its continued use and long-term survival.

This master thesis aims to explore and address the challenges associated with enhancing translation capabilities for low-resource languages, with a specific focus on the Sudanese dialect. By leveraging advancements in NMT and employing innovative techniques tailored to low-resource scenarios, we seek to develop effective translation models that can contribute to the accessibility, development, and preservation of the Sudanese dialect. Additionally, this research endeavors to shed light on the broader implications and socio-cultural significance of translation for low-resource languages, emphasizing the transformative power it holds in fostering mutual understanding and empowering marginalized communities.

The subsequent chapters of this thesis will delve into the existing literature on NMT and low-resource languages, investigate the specific challenges faced by the Sudanese dialect, present the methodologies employed for training translation models, and evaluate the performance of these models against human translators. The findings and insights gained from this study will provide valuable contributions to the fields of machine translation, linguistics, and language preservation, ultimately aiding the efforts to bridge the translation gap for low-resource languages and foster inclusive linguistic environments.

Through this research, we hope to lay the groundwork for future advancements in translation technologies, ensuring equitable access to information, facilitating cross-cultural understanding, and preserving the linguistic diversity that is essential for a globalized yet inclusive society.

2 Literature Review

2.1 Introduction

Machine translation, a rapidly developing field, has gained significant attention in recent years, particularly in the domain of Arabic to English translation and vice versa. Arabic, as the Qur'anic language, holds great importance for millions of people worldwide who seek to understand the Quran, the holy book of Muslims. Thus, the development of effective techniques and computer systems for high-quality translations becomes crucial. This literature review explores existing research in the field of Arabic machine translation, focusing on the specific context of the Sudanese dialect.

2.2 Challenges in Arabic Machine Translation

Arabic poses unique challenges for machine translation due to its complex morphology. Researchers such as Habash and Sadat [11] emphasize the significance of tokenization, the process of segmenting Arabic text into meaningful units, for accurate translation. While basic punctuation tokenization suffices in some cases, it may not be sufficient for syntactic analysis.

Efforts have been made to develop Arabic morphologies using different strategies. One-level rules have been employed by Attia [2] and Alansary et al. [1] to analyze Arabic at the stem level. Köpr and Miller [12] presented a powerful Arabic morphological analyzer using multiple levels of analysis. However, challenges remain, such as errors stemming from words not present in the lexicon or incorrect analysis for certain source language words.

Syntactic analysis of Arabic sentences is another crucial aspect. Žabokrtský and Smrž [20] developed a dependency grammar for Arabic, while Ditters [7] utilized the Definite Clause Grammar formalism to identify syntactic ambiguity in Arabic sentences. Attia [3] built an Arabic parser within the LFG framework, demonstrating its applicability to short sentences in the news domain. Spence and Christopher [17] improved Arabic parsing performance by cataloging parsing errors and addressing segmentation errors.

2.3 Arabic-to-English Machine Translation

Researchers have explored various approaches to Arabic-to-English machine translation. Salem et al. [16] proposed a machine translation system based on Role and Reference Grammar, but its limited lexicon hindered accurate translations. Bisazza and Federico [5] and Carpuat et al. [6]focused on reordering techniques to handle verb movements in Arabic sentences. Nguyen and Vogel [13] employed a context-dependent morphology preprocessing technique, while Yassine et al. [19] investigated the use of lexical and syntactic features for Arabic named entity recognition.

2.4 English-to-Arabic Machine Translation

Researchers have also addressed challenges in English-to-Arabic translation, including orthographical differences and syntactic reordering. Badr et al. [4] applied syntactic phrase reordering to enhance translation quality, and Elming and Habash [8] studied the impact of alignment methods on learning reordering rules. Toutanova et al. [18] integrated inflection generation models with statistical machine translation, achieving improved accuracy in translating English to morphologically complex languages like Arabic.

2.5 The Sudanese Dialect and Machine Translation

While the existing literature provides valuable insights into Arabic machine translation, there is a limited focus on the Sudanese dialect. The Sudanese dialect presents unique linguistic characteristics and challenges, such as dialectal variations and distinct word usage. Therefore, this project aims to address this gap by developing efficient techniques and systems specifically tailored to the Sudanese dialect. By considering the intricacies of this dialect, the project intends to contribute to the advancement of machine translation in the Sudanese context.

3 Methodology

This section presents the methodology employed in the thesis which focuses on the translation of a subset of 911 sentences from the FLORES101 English dataset [10] to the Sudanese dialect using the M2M100 multi-lingual many-to-many encoder-decoder machine translation model [9].

3.1 Dataset Collection

To facilitate the study on low-resource Sudanese dialect translation, a crucial step involved collecting a suitable dataset. In this section, we outline the process of obtaining the Sudanese dialect dataset, which involved translating a subset of 911 sentences from the Flores101 dataset. The collection process was made possible through the valuable contributions of multiple voluntary collaborators.

The Flores101 dataset, a multilingual parallel corpus, served as the foundation for acquiring a subset of sentences to represent the Sudanese dialect. This dataset covers a diverse range of languages, including English, providing a broad spectrum of translations to support the fine-tuning of translation models. Considering the scarcity of specific Sudanese dialect data, this approach enabled us to leverage existing resources and adapt them to the Sudanese dialect context.

To ensure the accuracy and quality of the translated dataset, a team of voluntary collaborators was formed. These collaborators were native speakers and possessed a deep understanding of the Sudanese dialect. Their expertise and linguistic proficiency were instrumental in capturing the nuances, expressions, and cultural references unique to the Sudanese dialect. The involvement of multiple collaborators added diversity and reduced individual biases, resulting in a more comprehensive representation of the Sudanese dialect in the translated dataset.

The dataset translation process was carried out collaboratively, with clear guidelines provided to ensure consistency and maintain the integrity of the Sudanese dialect. Collaborators worked diligently to translate the subset of 911 sentences from the Flores101 dataset, taking into account the linguistic idiosyncrasies specific to the Sudanese dialect. Any ambiguities or uncertainties encountered during the translation process were collectively discussed and resolved to ensure accurate representations of the original sentences.

It is important to note that the translation of the dataset was a voluntary effort and relied on the goodwill and dedication of the collaborators. Their commitment and linguistic expertise played a vital role in the successful collection of the Sudanese dialect dataset.

To enhance the accessibility and reproducibility of our work, the code and dataset utilized in this study have been made available on a GitHub repository. Interested readers can access the repository [1]to obtain further details about the dataset collection process and to review the fine-tuning code employed in this research.

3.2 Fine-tuning M2M100

To assess the effectiveness of the M2M100 model, several experiments were conducted using different train/test splits. Firstly, the model was fine-tuned on an English-Sudanese dataset. In one experiment, the entire collected dataset was used for both training and testing. Subsequently, a second experiment was performed with 455 sentences used for training and 456 sentences for testing. These experiments aimed to compare the model's performance under different training conditions and dataset sizes.

Following this, the M2M100 model was fine-tuned on a Sudanese-English dataset. Similar to the previous case, one experiment utilized the entire collected dataset for training and testing. In the second experiment, 455 sentences were selected for training, while 456 sentences were used for testing. This approach allowed for a comprehensive evaluation of the model's translation capabilities in both translation directions. Additionally, the M2M100 model was fine-tuned on the English-Arabic subset of the FLORES101 dataset that was collaboratively translated. Two directions of translation, English-Arabic and Arabic-English, were explored. One experiment involved using the entire dataset for both training and testing, while another experiment utilized 455 sentences for training and 456 sentences for testing. These experiments aimed to assess the model's performance on the Flores101 English-Arabic subset.

3.3 Prompting ChatGPT

To provide a comparative analysis, the performance of the fine-tuned M2M100 model was compared with the Chat GPT model [14]. Chat GPT was prompted to produce translations for both the English-Sudanese and Sudanese-English datasets. The following prompt was used to instruct Chat GPT to generate the translations from Sudanese to English

"follow these instructions strictly.
- translate these ten sentences from sudanese to english.
- the sentences are separated by new line.
- only output the translation.
- output a new line after each sentence translation.
sentence1
...
...
sentence10
"

Furthermore, Chat GPT was employed to generate translations for the English-Arabic version of the dataset. These translations were used as a benchmark to evaluate the performance of the fine-tuned M2M100 model.

[1]https://github.com/mohamedelabbas1996/Sudanese-dialect-machine-translation

The resulting translations from both the M2M100 model and Chat GPT were evaluated based on BLEU score [15] using different values for the number of n-grams. The analysis of this evaluation provided insights into the effectiveness of the translation models and their suitability for translating between English, Sudanese dialect. The findings from these experiments will be presented and discussed in subsequent chapters, contributing to a comprehensive understanding of the translation capabilities of the M2M100 model in comparison to the Chat GPT model.

4 Evaluauation and Results

This section presents the evaluation and results obtained from the translation experiments conducted using the M2M100 model and Chat GPT. The performance of the translation models was assessed using the BLEU score, which measures the similarity between the model translations and the reference translations. Different n-gram orders, including BLEU-1, BLEU-2, BLEU-3, and BLEU-4, were utilized to evaluate the quality of the translations.

Table 1: Evaluation Results

Model	Dataset	Train	Test	BLEU-1	BLEU-2	BLEU-3	BLEU-4
M2M100-fine-tuned	FLORES-101(English-Sudanese)	911	911	0.953	0.942	0.932	0.922
M2M100-fine-tuned	**FLORES-101(English-Sudanese)**	455	461	**0.315**	**0.198**	**0.128**	**0.084**
M2M100-fine-tuned	FLORES-101(Sudanese-English)	911	911	0.974	0.967	0.961	0.954
M2M100-fine-tuned	**FLORES-101(Sudanese-English)**	455	456	**0.467**	**0.322**	**0.230**	**0.168**
M2M100-fine-tuned	FLORES-101(English-Arabic)	911	911	0.971	0.964	0.958	0.951
M2M100-fine-tuned	FLORES-101(English-Arabic)	455	456	0.438	0.312	0.231	0.173
M2M100-fine-tuned	FLORES-101(Arabic-English)	455	456	0.501	0.366	0.273	0.208
M2M100-fine-tuned	FLORES-101(Arabic-English)	455	456	0.558	0.415	0.316	0.244
ChatGPT	FLORES-101(English-Sudanese)	0	911	0.309	0.196	0.129	0.087
ChatGPT	FLORES-101(Sudanese-English)	0	911	0.578	0.445	0.352	0.284
ChatGPT	FLORES-101(English-Arabic)	0	911	0.450	0.328	0.247	0.188

The evaluation results presented in Table 1 demonstrate the performance of the translation models across different n-gram orders.

The BLEU score results provide a quantitative measure of the translation quality for both models. However, it is important to consider other factors such as fluency, adequacy, and contextual understanding in assessing the overall translation performance. These additional aspects can be evaluated through human evaluation, where native speakers review and rate the translations based on these criteria.

The evaluation results demonstrate that both the M2M100 model and ChatGPT show promising translation capabilities. The M2M100 model exhibits higher BLEU scores across two n-gram orders, indicating a relatively higher level of similarity to the reference translations . However, further analysis considering other factors is necessary to provide a comprehensive assessment of the translation quality. The subsequent sections will delve into a detailed analysis of the translations and discuss the strengths and limitations of the models based on the evaluation results.

5 Discussion

5.1 Introduction

Translation between low-resource dialects poses a significant challenge due to the scarcity of available data for training translation models. In this discussion, we analyze the performance of two models, M2M100 fine-tuned model and ChatGPT, for translating the Sudanese dialect. Specifically, we examine the impact of fine-tuning on a limited dataset, the influence of n-grams on BLEU scores, and the translation quality in different language directions. Furthermore, we investigate the reasons behind potential mistranslations by ChatGPT in the context of Sudanese dialect. To compare the translation quality of the M2M100 fine-tuned model and ChatGPT, we present translation examples in Table 2 4 and Table 3 4,

Table 2: M2M100 fine-tuned Sample Translations

English	Sudanese (Reference)	Sudanese (Generated)
Ancient China had a unique way of showing different time periods; each stage of China or each family that was in power was a distinctive dynasty.	الصين القديمة كان عندها طريقة فريدة بتوضح بيها الفترات الزمنية مختلفة. كل مرحلة من مراحل الصين أو كل عائلة كانت في السلطة هي كانت سلالة حاكمة مميزة.	الصين القديمة كانت عندها طريقة فريدة بتظهر الفترات الزمنية مختلفة، كل مرحلة من مراحل الصين أو كل عائلة كانت في السلطة كانت سلالة مميزة.
Also between each dynasty was an unstable age of divided provinces. The best-known of these periods was the Three Kingdoms epoch taking place for 60 years between the Han and the Jin Dynasty.	برضو بين كل سلالة كان في عصر ما مستقر من المحافظات الاتقسمت. أشهر فترة من الفترات هي كانت حقبة الممالك التلاتة الاستمرت 60 سنة و كانت بين أسرة هان وجين.	كان بين كل سلالة عصر غير مستقر للمقاطعات المنقسمة. كان العصر المألوف في هذه الفترات هو عصر الممالك التلاتة الذي بتاخد 60 سنة بين سلالة هان و سلالة جين.
The Three Kingdoms was one of the bloodiest eras in Ancient China's history thousands of people died fighting to sit in the highest seat in the grand palace at Xi'an.	كانت الممالك الثلاث واحدة من أكتر العصور الدموية في تاريخ الصين القديمة ، لانو في آلاف من الناس ماتوا وهم يتقاتلوا عشان يجلسوا في أعلى مقعد في القصر الكبير في شيان.	كانت الممالك التلاتة واحدة من أكثر العصور دموية في تاريخ الصين القديمة آلاف الأشخاص ماتوا في معركة عشان يجلسوا في أعلى مقعد في القصر الكبير في شيان.
Regional and seasonal severe weather phenomena include blizzards, snowstorms, ice storms, and dust storms.	ظواهر الجو القاسي الإقليمية و الموسمية بتشمل العواصف التلجية العنيفة و العواصف التلجية والعواصف الجليدية و الكتّاحات.	الظواهر الجوية القاسية الإقليمية والموسمية بتشمل البرق والعواصف الثلجية وعواصف الجليد وعواصف الغبار.

respectively. Table 2 showcases the translations produced by the M2M100 fine-tuned model, highlighting its performance in capturing the nuances of the Sudanese dialect. Conversely, Table 3 demonstrates the translations generated by ChatGPT, shedding light on its strengths and potential limitations when translating the Sudanese dialect.

5.2 Performance of Fine-Tuned Models

When fine-tuning the M2M100 model on a limited dataset consisting of 455 sentences, the model exhibits promising performance and demonstrates good characteristics in capturing the nuances of the Sudanese dialect. This indicates that even with a small amount of data, the model can acquire a reasonable understanding of the target language.

5.3 Impact of n-grams on BLEU Scores

The evaluation of the translation quality using BLEU scores reveals a decrease in performance as the number of n-grams increases. This finding suggests that the models struggle to accurately capture the context and maintain coherence when translating longer sequences.

5.4 Directional Translation Quality

Comparing the translation quality between the Sudanese-English and English-Sudanese directions, it is observed that the M2M100 model performs better in translating Arabic-English than English-Arabic. This disparity may be attributed to the model's inherent strengths in handling Arabic-English translations, which are further supported by the availability of training data in this language pair.

Table 3: ChatGPT Sample Translations

English	Sudanese (Reference)	Sudanese (Generated)
Ancient China had a unique way of showing different time periods; each stage of China or each family that was in power was a distinctive dynasty.	الصين القديمة كان عندها طريقة فريدة بتوضح بيها الفترات الزمنية مختلفة. كل مرحلة من مراحل الصين أو كل عائلة كانت في السلطة هي كانت سلالة حاكمة مميزة.	الصين القديمة كانت لديها طريقة فريدة لإظهار الفترات الزمنية المختلفة؛ كانت كل مرحلة من مراحل الصين أو كل عائلة كانت في السلطة سلالة مميزة.
Also between each dynasty was an unstable age of divided provinces. The best-known of these periods was the Three Kingdoms epoch taking place for 60 years between the Han and the Jin Dynasty.	برضو بين كل سلالة كان في عصر ما مستقر من المحافظات الاتقسمت. أشهر فترة من الفترات هي كانت حقبة الممالك التلاتة الاستمرت 60 سنة و كانت بين أسرة هان وجين.	وكذلك بين كل سلالة كان هناك عصر غير مستقر من المقاطعات المنقسمة. وأشهر هذه الفترات الحقبة الثلاثية لمدة 60 عامًا بين ديناستي هان وجين.
The Three Kingdoms was one of the bloodiest eras in Ancient China's history thousands of people died fighting to sit in the highest seat in the grand palace at Xi'an.	كانت الممالك الثلاث واحدة من أكتر العصور الدموية في تاريخ الصين القديمة ، لانو في آلاف من الناس ماتوا وهم يتقاتلوا عشان يجلسوا في أعلى مقعد في القصر الكبير في شيان.	وكانت الحقبة الثلاثية واحدة من أقدم فترات القتال في تاريخ الصين القديمة حيث مات الآلاف من الناس يحاربون من أجل الجلوس في المقعد الأعلى في القصر العظيم في شيآن.
Regional and seasonal severe weather phenomena include blizzards, snowstorms, ice storms, and dust storms.	ظواهر الجو القاسي الإقليمية و الموسمية بتشمل العواصف التلجية العنيفة و العواصف التلجية والعواصف الجليدية و الكتاحات.	الظواهر الجوية الشديدة الإقليمية والموسمية تشمل العواصف الثلجية والعواصف الثلجية والعواصف الجليدية والعواصف الترابية.

5.5 Comparing M2M100 and ChatGPT

Regarding the performance comparison between M2M100 and ChatGPT for low-resource Sudanese dialect translation, the BLEU score analysis reveals that M2M100 outperforms ChatGPT when measuring the translation quality from English to Sudanese. This outcome is expected given the limited amount of data used to fine-tune the model. Conversely, ChatGPT demonstrates superior performance when translating from the Sudanese dialect to English, likely due to the extensive training data available during the training of ChatGPT.

5.6 Challenges in Sudanese Dialect Translation

One prominent challenge observed in ChatGPT's translation of Sudanese dialect sentences is the occasional mistranslation. This can be attributed to the unique expressions, idioms, and cultural references often present in the Sudanese dialect, which may not be widely known or understood outside of the local community. Consequently, without exposure to such specific regional knowledge, the model may produce inaccurate or nonsensical translations, thereby highlighting the importance of cultural and contextual understanding in dialect translation.

6 Conclusion

This research project aimed to enhance the accessibility of English content for speakers of the Sudanese dialect, a low-resource language, by fine-tuning an English-Arabic multilingual machine translation model. A dataset of 911 translated sentences in the Sudanese dialect was developed to train the model. Comparative analysis demonstrated the improved performance of the fine-tuned model compared to an advanced language model for low-resource Sudanese dialect translation. The findings highlight the potential of the model to address language resource disparities and its versatility for translation tasks. This research contributes to bridging the language gap and provides a foundation for further advancements in machine translation for the Sudanese dialect and other low-resource languages.

References

[1] Sherif Alansary, Magdy Nagi, and Nizar Adly. Towards analysing the international corpus of arabic (ica). In *International conference on language engineering. Progress of Morphological Stage*, pages 241–245, Egypt, 2009.

[2] Mohammed Attia. Developing a robust arabic morphological transducer using finite state technology. In *The 8th annual CLUK research colloquium*, Manchester, 2005.

[3] Mohammed Attia. *Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation.* PhD thesis, The University of Manchester, Manchester, 2008.

[4] Ibrahim Badr, Rabih Zbib, and James Glass. Syntactic phrase reordering for english-to-arabic statistical machine translation. In *The 12th conference of the European chapter of the association for computational linguistics*, pages 86–93, Athens, 2009.

[5] Arianna Bisazza and Marcello Federico. Chunk-based verb reordering in vso sentences for arabic-english statistical machine translation. In *ACL 2010: joint fifth workshop on statistical machine translation and MetricsMATR. Proceedings of the workshop*, pages 235–243, Uppsala University, Uppsala, July 2010.

[6] Marine Carpuat, Marton Yuval, and Habash Nizar. Improving arabic-to-english statistical machine translation by reordering post-verbal subjects for alignment. In *ACL 2010: the 48th annual meeting of the association for computational linguistics. Proceedings of the Conference Short Papers*, pages 178–183, Uppsala, July 2010.

[7] Elke Ditters. A formal grammar for the description of sentence structure in modern standard arabic. In *Workshop on Arabic processing: status and prospects at ACL/EACL*, Toulouse, 2001.

[8] Jakob Elming and Nizar Habash. Syntactic reordering for english-arabic phrase-based machine translation. In *Proceedings of the EACL 2009 workshop on computational approaches to semitic languages*, pages 69–77, Athens, 2009.

[9] 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. Beyond english-centric multilingual machine translation, 2020.

[10] Pedro M. García and Johannes Furnkranz. Flores101: A crowdsourced resource for low-resource machine translation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 4429–4438, 2021.

[11] Nizar Habash and Fatiha Sadat. Arabic pre-processing schemes for statistical machine translation. In *Proceedings of the 7th meeting of the North American chapter of the association for computational linguistics/human language technologies conference (HLT-NAACL06)*, pages 49–52, New York, 2006.

[12] Seniz Köpr and John Miller. A unification based approach to the morphological analysis and generation of arabic. In *CAASL-3—third workshop on computational approaches to Arabic script-based languages [at] MT Summit XII*, August 2009.

[13] Thanh Nguyen and Stephan Vogel. Context-based arabic morphological analysis for machine translation. In *Proceedings of the 12th conference on computational natural language learning*, pages 135–142, Manchester, 2008.

[14] OpenAI. Chatgpt (may 24 version) [large language model]. `https://chat.openai.com/`, 2023.

[15] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics (ACL)*, pages 311–318, 2002.

[16] Yasser Salem, Heinz Arnold, and Nolan Brian. Implementing arabic to english machine translation using the role and reference grammar linguistic model. In *Proceedings of the eighth annual international conference on information technology and telecommunication (ITT 2008)*, Galway, Ireland, October 2008. Runner-up for Best Paper Award.

[17] Graham Spence and Djon Christopher. Better arabic parsing: baselines, evaluations, and analysis. In *Coling 2010: 23rd international conference on computational linguistics. Proceedings of the conference*, pages 23–27, Beijing International Convention Centre, Beijing, August 2010.

[18] Kristina Toutanova, Hideki Suzuki, and Allen Ruopp. Applying morphology generation models to machine translation. In *ACL-08: HLT. 46th annual meeting of the association for computational linguistics: human language technologies. Proceedings of the conference*, pages 514–522, The Ohio State University, Columbus, June 2008.

[19] Bouziat Yassine, Zitouni Imed, Diab Mona, and Rosso Paolo. Arabic named entity recognition: using features extracted from noisy data. In *Proceedings of the ACL 2010 conference short papers*, pages 281–285, Uppsala, July 2010. c 2010 Association for Computational Linguistics.

[20] Zdeněk Žabokrtský and Otakar Smrž. Arabic syntactic trees: from constituency to dependency. In *The 10th conference of the European chapter of the association for computational linguistics*, pages 183–186, Budapest, 2009.