

Ministry of Higher Education and Scientific Research

وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba

Technology Faculty

Computer Sciences Department



جامعة باجي مختار – عنابة

كلية التكنولوجيا
قسم الإعلام الآلي

Thesis

Presented for obtaining diploma of

Doctorate

Specialty: Artificial Intelligence and Image Processing

Domain: Computer Sciences

By:

Amel Slim

Theme:

Artificial Intelligence Applied to Algerian Dialect Translation

Thesis defended in **December 13, 2023** In front of the committee:

N°	Full Name	Grade	Establishment	Quality
01	Hayet Farida Merouani	Prof.	Badji Mokhtar -Annaba University- Algeria	Chair
02	Ahlem Melouah	Prof.	Badji Mokhtar -Annaba University- Algeria	Rapporteur
03	Usef Faghihi	Prof.	Quebec, Trois-Rivières University- Canada	Co-rapporteur
04	Nabiha Azizi	Prof	Badji Mokhtar -Annaba University- Algeria	Examiner
05	Brahim Farou	Prof	University 08 Mai 1945-Guelma	Examiner



To my dear parents with all my love and for everything they have given me.

Acknowledgments

First and foremost, I express my heartfelt gratitude to **Allah**, the Almighty, who has bestowed abundant blessings, knowledge, and opportunities, enabling me to complete this thesis.

The fruition of this thesis would not have been possible without the unwavering support of numerous individuals. My sincere appreciation goes to my advisor, **Professor Ahlem Melouah**, whose guidance, advice, and mentorship have been invaluable throughout this journey. I extend my gratitude to my co-advisor, **Professor Usef Faghihi**, for being very generous with his time and knowledge and consistently assisting me at every stage of my Ph.D.

I extend my thanks to **Professor Hayet Farida Merouani** for graciously agreeing to preside over the thesis jury. Additionally, I express my gratitude to **Professor Nabiha Azizi** and **Professor Brahim Farou** for the honor of serving as members of my thesis jury.

Finally, I extend profound thanks to my parents, sisters, brothers, extended family, and numerous friends who stood by me throughout this extensive process, providing support and love. Their endurance and encouragement have been invaluable to me.

Abstract

The Algerian dialect is spoken by over 40 million people in the Algeria country. This dialect has developed over centuries, influenced by Arabic, French, and Berber languages. As a result, it has its own distinct vocabulary, grammar, and pronunciation, which makes it easier to use than the official language. Despite its unofficial status, the Algerian dialect is widely spoken and understood throughout the country, serving as a reminder of Algeria's complex linguistic heritage. Over time, this dialect has become an essential part of Algerian culture and identity, reflecting the rich history of immigration and cultural exchange that has shaped the region.

The Algerian dialect, despite its widespread use among Algerians and on social networking sites, can be incomprehensible to non-Algerians. The unique vocabulary, pronunciation, and grammar of this dialect make it challenging for those unfamiliar with it to understand. Therefore, translation services are necessary to ensure effective communication between Algerians and non-Algerians.

Neural machine translation (NMT) has emerged as a highly promising translation approach in recent years, attracting considerable interest. It relies on deep learning to generate translations and has consistently delivered remarkable outcomes across various languages. Its adaptability to language-specific features contributes to its success.

Our objective is to utilize NMT models, sequence-to-sequence(seq2seq), attentional sequence-to-sequence (Attentional seq2seq), and transformer for translating the Algerian dialect. Nevertheless, that dialect suffer from a scarcity of resources, and creating a database from scratch is a time-consuming and costly process.

To overcome the challenges posed by the limited resources available for Algerian dialect, we adopted the transfer learning strategy to enhance the performance of our NMT system. Transfer learning involves leveraging the knowledge gained from training a model on one task to improve the performance of a related but distinct task.

Our approach utilizes transfer learning to leverage the shared knowledge between Arabic dialects and Algerian dialect. Initially, the NMT model is trained on Arabic dialects and subsequently retrained on Algerian dialect. This approach yielded promising outcomes, with the BLEU score increasing from 0.5 to 34 and from 16.5 to 35 seq2seq, and Attentional seq2seq models respectively. Another experiment demonstrated a +1 points

blue score improvement with the attentional seq2seq and transformer models. To confirm the efficacy of this strategy, we also applied it to Moroccan and Tunisian dialects, obtaining favorable results.

Subsequently, we introduced the incremental transfer learning strategy, which transfers knowledge from Arabic dialects to Maghreb dialects and then to Algerian dialects. This strategy led to a +2 points blue score improvement with attentional seq2seq and +6 points blue score improvement with transformer for Algerian dialects. In addition, it led to a +2 points blue score improvement with transformer and a +1 point improvement with attentional seq2seq, for Moroccan and Tunisian dialects.

Key words: Machine Translation, Neural Machine Translation (NMT), Sequence-to-Sequence (seq2seq), Attention Mechanism, Transformer, Arabic Dialects, Algerian Dialect, Transfer Learning, Shared Features, Instance based Transfer Learning.

الملخص

اللهجة الجزائرية يتحدث بها أكثر من 40 مليون شخص في الجزائر. تطورت هذه اللهجة على مر القرون، متأثرة باللغات العربية والفرنسية والأمازيغية. نتيجة لذلك، لديها مفرداتها وقواعدها ونطقها المميز الذي يجعل استخدامها أسهل من اللغة الرسمية. على الرغم من وضعها غير الرسمي، إلا أن اللهجة الجزائرية متحدث بها ومفهومة على نطاق واسع في جميع أنحاء البلاد، مما يذكر بالتراث اللغوي المعقد للجزائر. مع مرور الوقت، أصبحت هذه اللهجة جزءاً أساسياً من الثقافة والهوية الجزائرية، عاكسة الثراء التاريخي للهجرة والتبادل الثقافي الذي شكل المنطقة.

إن اللهجة الجزائرية على الرغم من استخدامها الشائع بين الجزائريين وفي مواقع التواصل الاجتماعي، غير مفهومة لغير الجزائريين. المفردات المميزة والنطق والقواعد الخاصة بهذه اللهجة تجعلها تحدياً لأولئك الذين غير ملمين بها لفهمها. لذلك، فإن خدمات الترجمة ضرورية لضمان التواصل الفعال بين الجزائريين وغير الجزائريين.

ظهرت الترجمة الآلية العصبية كنهج ترجمة مثير للأمل في السنوات الأخيرة وحظيت باهتمام كبير. تعتمد على التعلم العميق لإنتاج الترجمات وقد أسفرت دائماً عن نتائج ملحوظة في مختلف اللغات وقابلية التكيف مع الميزات الخاصة بكل لغة تسهم في نجاحها.

هدفنا هو استخدام نماذج NMT، التسلسل إلى التسلسل (seq2seq)، التسلسل إلى التسلسل المتعمد على خاصية الانتباه (Attentional seq2seq)، والمحول (Transformer) لترجمة اللهجة الجزائرية. ومع ذلك، فإن هذه اللهجة تعاني من ندرة الموارد، ويعد إنشاء قاعدة بيانات من الصفر عملية مكلفة وتستغرق وقتاً طويلاً.

للتغلب على التحديات التي تفرضها محدودية الموارد المتاحة للهجة الجزائرية، اعتمدنا استراتيجية التعلم الانتقالي لتعزيز أداء نظام NMT الخاص بنا. يتضمن التعلم الانتقالي الاستفادة من المعرفة المكتسبة من تدريب نموذج على مهمة واحدة لتحسين أداء مهمة ذات صلة ولكنها مميزة.

يستخدم نهجنا التعلم الانتقالي للاستفادة من المعرفة المشتركة بين اللهجات العربية واللهجة الجزائرية. في البداية، تم تدريب نموذج NMT على اللهجات العربية ثم أعيد تدريبه على اللهجة الجزائرية. أسفر هذا النهج عن نتائج واعدة، حيث زادت درجة BLEU من 0.5 إلى 34 ومن 16.5 إلى 35 seq2seq ونماذج Attentional seq2seq على التوالي. أظهرت تجربة أخرى تحسناً في درجة BLEU بمقدار +1 نقطة باستخدام نماذج Attentional seq2seq و Transformer. لتأكيد فاعلية هذه الاستراتيجية، قمنا بتطبيقها أيضاً على اللهجات المغربية والتونسية، وحصلنا على نتائج إيجابية.

بعد ذلك، قدمنا استراتيجية التعلم التحويلية التدريجية، والتي تنقل المعرفة من اللهجات العربية إلى اللهجات المغاربية ثم إلى اللهجات الجزائرية. أدت هذه الإستراتيجية إلى تحسين درجة BLEU بمقدار +2 نقطة مع Attentional seq2seq وبمقدار +6 نقاط باستخدام Transformer للهجات الجزائرية. بالإضافة إلى ذلك، أدى ذلك إلى تحسين درجة BLEU بمقدار +2 نقطة مع المحول وتحسين +1 نقطة مع Attentional seq2seq للهجات المغربية والتونسية.

الكلمات الرئيسية: الترجمة الآلية، الترجمة الآلية العصبية (NMT) ، تسلسل-إلى-تسلسل (seq2seq) ، آلية الانتباه، المحوّل (Transformer) ، اللهجات العربية، اللهجة الجزائرية، نقل المعرفة، الميزات المشتركة، التعلم بالنقل المستند إلى المثل.

Résumé

Le dialecte algérien est parlé par plus de 40 millions de personnes en Algérie. Ce dialecte s'est développé au fil des siècles, influencé par l'arabe, le français et les langues berbères. Il possède son propre vocabulaire, sa propre grammaire et sa propre prononciation, ce qui le rend plus facile à utiliser que la langue officielle. Malgré son statut non officiel, le dialecte algérien est largement parlé et compris dans tout le pays, rappelant l'héritage linguistique complexe de l'Algérie. Au fil du temps, ce dialecte est devenu un élément essentiel de la culture et de l'identité algériennes, reflétant la riche histoire de l'immigration et des échanges culturels qui ont façonné la région.

Le dialecte algérien, bien qu'il soit largement utilisé par les Algériens et sur les sites de réseautage social, peut être incompréhensible pour les non-algériens. Le vocabulaire, la prononciation et la grammaire propres à ce dialecte rendent sa compréhension difficile pour ceux qui ne le connaissent pas. Les services de traduction sont donc nécessaires pour assurer une communication efficace entre Algériens et non-Algériens.

La traduction automatique neuronale (NMT) est apparue comme une approche de traduction très prometteuse ces dernières années, suscitant un intérêt considérable. Elle s'appuie sur l'apprentissage profond pour générer des traductions et a toujours produit des résultats remarquables dans différentes langues. Son adaptabilité aux caractéristiques propres à chaque langue contribue à son succès.

Notre objectif est d'utiliser des modèles de NMT, séquence à séquence (seq2seq), séquence à séquence basé sur le mécanisme d'attention (Attentional seq2seq), et le Transformer pour traduire le dialecte algérien. Néanmoins, ce dialecte souffre d'une pénurie de ressources, et la création d'une base de données à partir de zéro est un processus long et coûteux.

Pour surmonter les défis posés par les ressources limitées disponibles pour le dialecte algérien, nous avons adopté la stratégie de l'apprentissage par transfert afin d'améliorer les performances de notre système de NMT. L'apprentissage par transfert consiste à tirer parti des caractéristiques acquises lors de l'entraînement d'un modèle à un domaine pour améliorer les performances d'un domaine connexe mais distincte.

Notre approche utilise l'apprentissage par transfert pour exploiter les caractéristiques partagées entre les dialectes arabes et le dialecte algérien. Dans un premier temps, le mod-

èle NMT est entraîné sur les dialectes arabes, puis réentraîné sur les dialectes algériens. Cette approche a donné des résultats prometteurs, le score BLEU passant respectivement de 0,5 à 34 et de 16,5 à 35 pour les modèles seq2seq et attentionnel seq2seq. Une autre expérience a démontré une amélioration du score bleu de +1 point avec les modèles attentionnel seq2seq et Transformer. Pour confirmer l'efficacité de cette stratégie, nous l'avons également appliquée aux dialectes marocains et tunisiens, obtenant des résultats favorables.

Par la suite, nous avons introduit la stratégie d'apprentissage par transfert incrémental, qui transfère les caractéristique des dialectes arabes aux dialectes maghrébins, puis aux dialectes algériens. Cette stratégie a conduit à une amélioration du score bleu de +2 points avec l'attentionnel seq2seq et de +6 points avec le Transformer pour les dialectes algériens. En outre, elle a conduit à une amélioration du score bleu de +2 points avec le Transformer et de +1 point avec attentionnel seq2seq, pour les dialectes marocains et tunisiens.

Mots clés : Traduction automatique, traduction automatique neuronale (NMT), séquence à séquence (seq2seq), mécanisme d'attention, Transformer, dialectes arabes, dialecte algérien, apprentissage par transfert, caractéristiques partagées, apprentissage par transfert basé sur l'instance.

List of Abbreviation

MT	Machine Translation
NMT	Neural Machine Translation
NLP	Natural Language Processing
Seq2seq	Sequence-to-Sequence
Attentional seq2seq	Attentional Sequence-to-Sequence
MSA	Modern Standart Arabic
SL	Source Language
TL	Target Language
RBMT	Rule-based Machine Translation
CBMT	Corpus-based Machine Translation
SMT	Statistical Machine Translation
EBMT	Example-based Machine Translation
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory
GRU	Gated Recurrent Unit
CNN	Convolutionnal Neural Network
SVO	Subject-Verb-Object
SOV	Subject-Object-Verb
VSO	Subject-Object-Verb
OSV	Object-Subject-Verb
VOS	Verb-Object-Subject
VSO	Verb-Subject-Object

OOV	Out-of-Vocabulary
MADAR	Multilingual Arabic Dialect Applications and Resources
PADIC	Arabic Parallel Dialects Corpus
MPCA	Multidialectal Parallel Corpus of Arabic
GNMT	Google’s Neural Machine Translation
ConvS2S	Convolutional Sequence-to-Sequence
BLEU	Bilingual Evaluation Understudy
METEOR	Metric for Evaluation of Translation with Explicit Ordering

Contents

Acknowledgments	iii
Abstract	iv
List of Abbreviation	x
List of Figures	xvi
List of Tables	xix
1 Introduction	1
1.1 Background	1
1.2 Problem statement	2
1.3 Purpose of the study and Contributions	4
1.4 Outline of the Thesis	5
2 Algerian Dialect : A Linguistic Overview	9
2.1 Introduction	9
2.2 Lexicology	9
2.2.1 Vocabulary	10
2.2.2 Semantics	10
2.2.3 Lexical Categories	11
2.3 Phonology	12
2.3.1 Phonetic Inventory	12
2.3.2 Phonological Process	14
2.3.3 Pronunciation	15
2.4 Morphology	15
2.4.1 Word Formation	16
2.4.2 Pronouns	16
2.4.2.1 Personal Pronouns	16
2.4.2.2 Possessive and Adjective Pronouns	17

2.4.2.3	Demonstrative Pronouns	18
2.4.3	Verb Conjugation	20
2.4.3.1	Past Tense	20
2.4.3.2	Present Tense	22
2.4.3.3	Future Tense	23
2.4.3.4	Imperative Mood	24
2.4.4	Word Declension	24
2.5	Syntax	26
2.5.1	Word Order	26
2.5.2	Clause Structure	27
2.5.3	Sentence Type	27
2.6	Conclusion	30
3	Machine Translation	31
3.1	Introduction	31
3.2	What's Machine Translation	31
3.3	Rule Based Machine Translation	32
3.3.1	Direct Machine Translation	32
3.3.2	Transfer Machine Translation	33
3.3.3	Interlingua Machine Translation	34
3.4	Corpus Based Machine Translation	34
3.4.1	Statistical Machine Translation	35
3.4.2	Example Based Machine Translation	35
3.4.3	Neural Machine Translation	36
3.4.3.1	Recurrent Neural Network	37
3.4.3.2	Word Embeddings	41
3.4.3.3	Sequence-to-Sequence	41
3.4.3.4	Attention Mechanism	44
3.4.3.5	Fundamental Architecture in NMT	46
3.5	Metric Evaluation of Machine Translation	53
3.5.1	BLEU Score	53
3.5.2	METEOR Score	54
3.6	Conclusion	55
4	Arabic Dialect Neural Machine Translation Overview and Contribution	56
4.1	Introduction	56
4.2	Arabic Dialects Parallel Corpora	57
4.3	Arabic Dialects NMT an Overview	59
4.4	NMT Models for Arabic Dialects a Contribution	61

4.4.1	Arabic Dialects Translation Using Seq2seq	63
4.4.1.1	Experimentation	65
4.4.1.2	Result	67
4.4.2	Arabic Dialects Translation Using Attentional Seq2seq	80
4.4.2.1	Experimentation	83
4.4.2.2	Result	83
4.4.3	Arabic Dialects Translation Using Transformer	100
4.4.3.1	Exprimentation	101
4.4.3.2	Result	102
4.4.4	Discussion	110
4.5	Conclusion	116
5	Transfer Learning Neural Machine Translation Overview and Contribution	117
5.1	Introduction	117
5.2	Transfer Learning	118
5.2.1	Transfer Learning Setting Categorization	120
5.2.1.1	Inductive Transfer Learning	120
5.2.1.2	Transductive Transfer Learning	121
5.2.1.3	Unsupervised Transfer Learning	121
5.2.2	Approaches to Transfer Learning Setting	121
5.2.2.1	Instance Based Transfer Learning	122
5.2.2.2	Feature Based Transfer Learning	124
5.2.2.3	Parameter Based Transfer Learning	125
5.2.2.4	Relational Knowledge Based Transfer Learning	126
5.3	Arabic Transfer learning NMT: an Overview	127
5.4	Algerian Dialect Transfer Learning NMT: a Contribution	129
5.4.1	Transductive Transfer Learning	130
5.4.1.1	Linguistic Background	131
5.4.1.2	Method	133
5.4.1.3	Experimentation	135
5.4.2	Incremental Transfer Learning: Proposed Strategy	137
5.4.2.1	Linguistic Background	137
5.4.2.2	Method	140
5.4.2.3	Experimentation	142
5.4.3	Result	143
5.4.4	Discussion	171
5.5	Conclusion	178
6	Conclusion	179

References	185
-------------------	------------

List of Publications	198
-----------------------------	------------

List of Figures

1.1	Thesis Map.	8
2.1	Difference between Noun Declension in Algerian Dialect and MSA. . . .	25
3.1	The Bernard Vauquois Pyramid of RBMT Approaches	33
3.2	Illustrating the Functioning of RNNs for Processing Sequential Data . . .	38
3.3	Seq2seq NMT Model Architecture.	42
3.4	Decoder Architecture in Seq2Seq NMT Model: Utilizing Context Vector for Output Prediction.	43
3.5	Bidirectional Seq2Seq NMT Model Architecture.	43
3.6	Stacked LSTM layers in Seq2seq NMT Model Architecture with Residual connections.	44
3.7	The GNMT Architecture.	46
3.8	The Convs2s Architecture.	48
3.9	The Transformer Architecture.	49
3.10	Attention Mechanisms in the Transformer. (a) Multi-Head Attention consists of several attention layers running in parallel, and (b) Scaled Dot-Product Attention.	51
3.11	Model architecture of RNMT+.	52
4.1	Number of Algerian Dialect Sentences in Existing Arabic Dialects parallel corpora. The figure highlights the scarcity of comprehensive parallel corpora containing Algerian dialect sentences among existing Arabic dialects corpora.	58
4.2	The NMT system's architecture employed for translating Arabic dialects to MSA.	62
4.3	The Architecture of seq2seq NMT Model used to translate the Arabic Dialect. 64	
4.4	The Architecture of Attentional seq2seq NMT Model used to translate the Arabic Dialect.	81

4.5	Comparison of BLEU score results for each experiment with different models: Seq2Seq, Attentional Seq2Seq, and Transformer. The figure showcases the performance of each model in Arabic dialect NMT tasks, highlighting their effectiveness in terms of BLEU scores.	111
4.6	BLEU score value relative to the length of the sentences for each experiment.	113
4.7	Comparison of Results from our Work and Related Studies.	115
5.1	The suitable approaches for transfer learning settings.	122
5.2	Relational Information Transfer from the Academic to Cinematic Domain. This figure illustrates the process of transferring relational information from the academic domain to the cinematic domain. The left side shows a knowledge graph representing academic concepts, while the right side shows a knowledge graph representing cinematic concepts. The arrows between the two graphs indicate the transfer of relational information using a transfer learning approach. The approach allows the model to leverage the knowledge learned in one domain to improve performance in another domain.	126
5.3	Heatmap of Similarity of some words between Algerian and Other Arabic Dialects. The heatmap depicts shared vocabulary between Algerian dialect and other Arabic dialects, using green to indicate similarity and yellow to indicate dissimilarity. Constructed using MADAR and PADIC datasets, the results demonstrate that the Algerian dialect shares many words with other Arabic dialects, especially those spoken in North Africa and the Middle East.	132
5.4	Lexical Dissimilarity between MADAR Arabic Dialects and Algerian Dialect from PADIC. This figure provides a visual representation of the dissimilarity at the lexical level between the Arabic dialects from the MADAR dataset (in blue) and the Algerian dialect from the PADIC dataset (in red), with the overlap of words between the two datasets in the middle. This figure emphasizes the importance of considering lexical differences when analyzing Arabic dialects.	134
5.5	The Architecture of the transductive transfer learning strategy used. . . .	135
5.6	The Architecture of the Incremental transfer learning strategy used. . . .	141
5.7	Unveiling the impact of Transductive Transfer Learning: A Comparative Analysis across Four Experiments with Seq2Seq, Attentional Seq2Seq, and Transformer Models.	172
5.8	Unveiling the impact of Incremental Transfer Learning: A Comparative Analysis across three Experiments with Attentional Seq2Seq and Transformer Models.	174

5.9 BLEU score balue relative to the length of the for each Transfer learning experiment.	177
---	-----

List of Tables

2.1	Loanwords in Algerian Dialect Vocabulary	11
2.2	Phonetic table of Algerian Dialect.	13
2.3	Differences in word pronunciation exist between the Algerian dialect. . .	15
2.4	Personal pernouns in Algerian Dialect and MSA.	17
2.5	Possissive pernouns in Algerian Dialect and MSA.	18
2.6	Possissive adjective in Algerian Dialect and MSA.	19
2.7	Proximal and Distal Demonstrative Pronouns in Algerian Dialect and MSA.	19
2.8	Past tense Verb conjugation in Algerian Dialect and MSA.	21
2.9	Present tense verb conjugation in Algerian Dialect and MSA.	21
2.10	Future tense verb conjugation in Algerian Dialect and MSA.	23
2.11	Imperative Mood verb conjugation in Algerian Dialect and MSA.	24
2.12	Interrogative Particles used in MSA, ALG and ANB Algerian Dialects. .	28
4.1	Related works of Arabic dialect NMT models.	60
4.2	BLEU scores results of four experiments evaluating the effectiveness of Seq2Seq model in Arabic Dialects translation.	67
4.3	Examples of Correct Translation Predictions by the Seq2seq Model in Ex- periment 1.	70
4.4	Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 1.	71
4.5	Examples of Correct Translation Predictions by the Seq2seq Model in Ex- periment 2.	72
4.6	Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 2.	73
4.7	Examples of Correct Translation Predictions by the Seq2seq Model in Ex- periment 3.	75
4.8	Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 3.	76
4.9	Examples of Correct Translation Predictions by the Seq2seq Model in Ex- periment 4.	78

4.10	Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 4.	79
4.11	BLEU scores results of seven experiments evaluating the effectiveness of Attentional Seq2Seq model in Arabic dialects translation.	84
4.12	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 2.	87
4.13	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 2.	88
4.14	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 3.	89
4.15	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 3.	90
4.16	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 4.	92
4.17	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 4.	93
4.18	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 5.	94
4.19	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 5.	95
4.20	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 6.	96
4.21	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 6.	97
4.22	Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 7.	98
4.23	Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 7.	99
4.24	BLEU scores results of three experiments evaluating the effectiveness of the Transformer model in Arabic dialects translation.	102
4.25	Examples of Correct Translation Predictions by the Transformer Model in Experiment 5.	104
4.26	Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 5.	105
4.27	Examples of Correct Translation Predictions by the Transformer Model in Experiment 6.	106
4.28	Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 6.	107

4.29	Examples of Correct Translation Predictions by the Transformer Model in Experiment 7.	108
4.30	Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 7.	109
5.1	Related works of Transfer Learning strategy applied on Arabic dialect NMT.	128
5.2	BLEU scores results of conventional Transfer learning experiments evaluating the effectiveness of the seq2seq model for Algerian Dialect translation.	144
5.3	BLEU scores results of incremental and conventional Transfer learning experiments evaluating the effectiveness of Attentional Seq2Seq model in Arabic dialects translation.	144
5.4	BLEU scores results of incremental and conventional Transfer learning experiments evaluating the effectiveness of the Transformer model in Arabic dialects translation.	145
5.5	Examples of Correct Translations Generated by Seq2Seq Model with Transfer Learning in Experiment 2-TL.	146
5.6	Examples of Partially Correct Translations Generated by Seq2Seq Model with Transfer Learning in Experiment 2-TL.	147
5.7	Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 2-TL.	149
5.8	Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 2-TL.	150
5.9	Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 5-TL.	151
5.10	Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 5-TL.	152
5.11	Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 6-TL.	153
5.12	Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 6-TL.	154
5.13	Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 7-TL.	156
5.14	Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 7-TL.	157
5.15	Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 5-TL.	158
5.16	Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 5-TL.	159

5.17	Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 6-TL.	161
5.18	Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 6-TL.	162
5.19	Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 7-TL.	163
5.20	Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 7-TL.	164
5.21	Examples of Correct Translations Generated by Attentional Seq2Seq Model with incremental transfer learning in Experiment 5-ITL.	165
5.22	Examples of Correct Translations Generated by Attentional Seq2Seq Model with incremental transfer learning in Experiment 6-ITL.	166
5.23	Examples of Correct Translations Generated by Attentional Seq2Seq Model with incremental transfer learning in Experiment 7-ITL.	167
5.24	Examples of Correct Translations Generated by Transformer Model with incremental transfer learning in Experiment 5-ITL.	168
5.25	Examples of Correct Translations Generated by Transformer Model with incremental transfer learning in Experiment 6-ITL.	169
5.26	Examples of Correct Translations Generated by Transformer Model with incremental transfer learning in Experiment 7-ITL.	170

1

Introduction

1.1 Background

Communication is the process of communicating ideas, opinions, facts, beliefs, and sentiments through a range of verbal and non-verbal message. Talking face to face, by phone, or by writing via social networking sites are the most used mediums in communication. Typically, messages in communication are conveyed in a language that is common to both the sender and the recipient. However, in some situations, the message may need to be translated from the sender's language into the recipient's language in order to ensure clear understanding. This translation process helps to bridge any linguistic barriers and ensures the effective transmission of the message.

In the world, there are many languages and dialects. Generally, a language is written as well as spoken, while a dialect is just spoken. It is not possible for everyone to learn all the languages and dialects present in the world to communicate, which is why translation is crucial for effective communication. Translation allows individuals who speak different languages or dialects to understand each other and exchange ideas, opinions, and information, bridging linguistic.

Arabic is an important language spoken in a wide geographic area, ranging from Mauritania in the west to Iraq in the east, where eighteen countries have it as their official language. Moreover, Arabic is also spoken by certain communities in Northern Cyprus, some areas of southern Turkey, and sub-Saharan Africa to the south. Additionally, Arabic-speaking communities still exist in the Balkh area of Afghanistan, sections of Iran, includ-

ing Khurasan in the east and Khuzistan in the south, as well as in Uzbekistan [1].

The Arabic language of the Middle East and North Africa has two primary forms: Modern Standard Arabic (MSA) and Arabic dialects. MSA is the formal, standardized version of the language that is used in writing, education, media, and formal speeches across the Arab world. It is also the version of Arabic used in international diplomacy, religious texts, and in some cases, news broadcasts. In contrast, Arabic dialects refer to the spoken forms of Arabic that differ in vocabulary, pronunciation, and grammar from MSA.

There are many different Arabic dialects spoken across the Arab world, with variations between countries, regions, and even between neighboring villages. Each dialect has its unique characteristics, reflecting the cultural and linguistic diversity of the Arabic-speaking population. For example, Algerian Dialect is one of the most widely spoken dialects of Arabic and is known for its distinct pronunciation, vocabulary, and grammar. While MSA is crucial for formal contexts and standardized communication, the various dialects are essential for everyday communication, allowing people to express themselves naturally in their local community and connect with others who speak the same dialect.

Effective communication is critical for promoting cultural understanding and collaboration between communities that speak different languages. However, the lack of standardization of Arabic dialects presents a unique challenge for communication between Arabic-speaking communities and other language-speaking communities. As a result, there is an urgent need for more accurate and effective translation systems to facilitate communication between Arabic-speaking and other language-speaking communities. This thesis aims to address the challenges of translating Algerian Dialect using machine translation (MT) technique, with the goal of contributing to the development of more reliable and accurate translation systems that can help overcome communication barriers within and between Arabic-speaking communities as well as between Arabic-speaking and other language-speaking communities.

1.2 Problem statement

Translating between Arabic dialects or from dialects to MSA or other language poses significant challenges for MT tools, mainly due to the lack of standardization and the vast linguistic diversity of the dialects. While some dialects may share similarities, there are often considerable differences in vocabulary, grammar, and pronunciation, which pose obstacles to developing accurate and reliable translation systems. For instance, translating from Algerian Dialect to MSA requires a deep understanding of the unique features of Algerian Dialect, such as its frequent use of French loanwords, and the ability to adapt to the syntax and grammar of MSA.

Furthermore, it's important to note that the Algerian dialect is not a singular, homogeneous dialect. Rather, it is a collection of sub-dialects that vary depending on the specific

region in Algeria. To illustrate, there are sub-dialects known as the ALG-Algerian dialect and the ANB-Algerian dialect. The ANB-Algerian dialect is specific to the Annaba region, whereas the ALG dialect refers to the dialect spoken in Algiers. Each sub-dialect has its own unique characteristics, including differences in vocabulary, grammar, and pronunciation. Therefore, translating different sub-dialects of the Algerian dialect into MSA or another language also presents its own set of challenges, such as understanding and accounting for the specific linguistic nuances of each sub-dialect. These complexities further highlight the need for translation systems that can accurately capture the nuances and diversity of the Algerian dialect.

Translating Algerian Dialect poses unique challenges for MT, as it is a highly complex dialect with significant regional variations and a lack of standardization. Rule-based machine translation (RBMT) and statistical machine translation (SMT) have proven to be somewhat effective in translating between Algerian Arabic, other Arabic dialects, MSA, and English [2–8], due to their ability to better manage the varied vocabulary and syntax found in these dialects. Nevertheless, RBMT and SMT still faces challenges in accurately capturing the nuances and cultural context of the dialect. NMT is a more recent approach that has shown promising results in improving the accuracy and naturalness of translations for complex languages. By leveraging deep learning algorithms and large amounts of training data, NMT can better handle the idiosyncrasies of the dialect and produce more natural-sounding translations. Despite its potential, NMT still faces challenges in terms of data availability and computational resources, but it represents a promising avenue for advancing the state of MT for Algerian dialect.

Furthermore, the lack of data is another significant challenge in translating Algerian dialect. Because dialect is often spoken rather than written, there is a shortage of high-quality data available to train NMT models. This shortage of data makes it difficult for NMT systems to accurately capture the nuances and subtleties of the Algerian dialect and can lead to errors and inaccuracies in translations. To overcome this challenge, researchers often need to collect and compile their own data sets, which can be time-consuming and resource-intensive. Additionally, the data sets may not be comprehensive enough to cover all the variations and sub-dialects of Algerian dialect, which can further limit the accuracy of MT systems.

The challenges in developing an effective MT system for Algerian dialect highlight the need for systems that can capture the complexity and diversity of the language, particularly when dealing with poorly-resourced dialects. In this context, this thesis seeks to explore the feasibility of using NMT models to tackle these challenges and create more reliable and accurate translation systems for Algerian Dialect. By addressing the challenges of translating between Algerian Dialect and MSA, this thesis aims to contribute to overcoming the communication barriers that exist within and between Algerian-speaking communities and other languages-speaking communities.

1.3 Purpose of the study and Contributions

As part of our investigation, we will aim to translate low-resource Algerian dialect without relying on data augmentation. To achieve this, we employed transfer learning in NMT. Transfer learning involves using knowledge learned from one task to improve performance on a different but related task. By leveraging transfer learning, we will aim to improve the accuracy and fluency of the translations without relying on additional training data.

Our transfer learning approach will implement in two stages. In the first stage, the model will train on the Arabic dialects to capture the commonalities and generalize the language structures. This step helps the model learn the common patterns that exist between the dialects and lay the foundation for better translation accuracy. In the second stage, the model will retrain on Algerian dialect, which enabled it to capture the unique features specific to this dialect, such as pronunciation and word usage. Overall, by leveraging existing data from similar dialects, our approach can help overcome the scarcity of training data and enhance the quality of translations.

Arabic dialects are typically grouped into five distinct categories based on the degree of similarity between the dialects [9–11]:

- The Maghrebi dialects are spoken in North Africa, including Algeria, Tunisia, Morocco, and Libya, and are characterized by the influence of Berber languages and a unique set of vocabulary and syntax.
- The Nile Basin dialects are spoken in Egypt, and Sudan, and are known for their distinct pronunciation and vocabulary.
- The Levantine dialects are spoken in Lebanon, Syria, Jordan, and Palestine, and have their own set of unique features, including distinct vocabulary and pronunciation.
- The Gulf dialects are spoken in the Arabian Gulf countries, including Saudi Arabia, Bahrain, Kuwait, Oman, Qatar, and the United Arab Emirates, and have a number of shared features.
- The Yemeni dialects are spoken in Yemen and are known for their distinct vocabulary and phonology.

This categorization is based on the geographic, historical, and cultural factors that have influenced the development of the various dialects over time. Furthermore, within each category, there exists a significant degree of convergence among the dialects in terms of lexicology, phonology, morphology, and syntax.

Within this framework of dialect categorization, the Algerian dialect can be categorized as part of the Maghrebi dialects, which, in turn, are classified under the broader category of Arabic dialects. This hierarchical structure highlights the regional association and shared

linguistic traits among the Maghrebi dialects, including Algerian dialect while acknowledging their overarching connection to the larger family of Arabic dialects. Each specific category inherits certain linguistic features from the broader category it belongs to. For instance, the Algerian dialect exhibits shared linguistic characteristics with the Maghrebi dialects, which in turn share linguistic features with the broader category of Arabic dialects. This hierarchical relationship allows for the transmission and continuity of linguistic traits, highlighting the interconnectedness and influence of the larger category on its respective subcategories.

It occurred to us that we could take advantage of these common linguistic features, instead of the model focusing on all the linguistic features, it can only focus on learning those that are very specific to the Algerian dialect. We sought to build on previous research and leverage the categorization of Arabic dialects into distinct groups. With the goal of developing an effective translation system, we will propose an incremental transfer learning approach, which involves training the model on progressively more specialized sets of data.

The first step of the incremental transfer learning approach will involve training the model on Arabic dialects. This allowed the model to learn the commonalities and shared features of Arabic dialects, which will be applied to other dialects within the language family. In the next stage, the model will retrain on Maghrebi dialects, by focusing on the features and nuances specific to the Maghrebi dialects, the model will be able to further improve its ability to translate these dialects.

Finally, in the last stage, the model will retrain on Algerian dialect, with the goal of achieving high accuracy and capturing the unique features of this low-resource language. This process of incremental transfer learning allows the model to gradually specialize on the target dialect, while building on the knowledge gained from previous stages. To implement our incremental transfer learning approach, we chose to use seq2seq, attentional seq2seq, and transformer models.

Interestingly, to validate the effectiveness of the strategies and models we will employ in our thesis, we will also extend their application to address the challenges posed by low-resource Tunisian and Moroccan dialects.

1.4 Outline of the Thesis

This thesis comprises six chapters. Chapter one serves as the general introduction, while chapters two and three discuss the theoretical background and foundations of the thesis statement. Chapters four and five present an overview of the thesis and its main contributions. Chapter six presents as a conclusion. The overall organization of the thesis is as follows:

Chapter two, titled "Algerian Dialect: A Linguistic Overview," provides an in-depth understanding of the Algerian dialect. This chapter is significant because it sets the

foundation for the entire thesis and helps to comprehend the rest of the research. In this chapter, we give a comprehensive linguistic overview of the dialect. We cover the essential elements of dialect, including lexicology, phonology, morphology, and syntax, which are all fundamental components of language. In the lexicology section, we present a detailed explanation of the Algerian dialect's vocabulary, semantics, and lexical categories. We delve into the different lexical categories, such as nouns, verbs, and pronouns, and explain how they operate in the Algerian dialect. We also highlight some unique features of the Algerian dialect that distinguish it from MSA. In the phonology section, we present the phonetic inventory, phonological processes, and pronunciation of words in the Algerian dialect. This section is essential because it helps to understand the sound system of the Algerian dialect and how it differs from MSA. In the morphology section, we explore the word formation process, including the formation of new words and the different types of pronouns (personal, possessive, adjective, and demonstrative pronouns). We also discuss verb conjugation (past, present, future tenses, and imperative mood), and word declension, which is an essential feature of the Algerian dialect. Lastly, we discuss syntax and its significance in the Algerian dialect. We examine word order, clause structure, and sentence types and explain how they influence the meaning of sentences in the dialect. Overall, chapter two serves as a critical component of the thesis, providing a detailed and comprehensive overview of the Algerian dialect's linguistic elements. The information presented in this chapter sets the stage for the rest of the thesis and helps to understand the research's significance.

In the third chapter of our thesis, titled "Machine Translation", we delve into the complex field of MT. We begin by introducing the concept of MT and discussing the two primary approaches: RBMT and CBMT. RBMT is the oldest approach, relying on human knowledge and a set of specific rules to translate text from one language to another. In this section, we delve deeper into RBMT, exploring the concept of sub-methods such as direct, transfer, and interlingua MT methods. Next, we move on to CBMT, which is based on the use of large data sets or corpora to train translation algorithms. Within CBMT, we present the concepts of SMT, EBMT, and NMT, which use different methods to derive translations from the training data. Focusing on NMT, which has become increasingly popular in recent years, we present seq2seq model, which is the foundation of NMT. We also discuss the key components of NMT, including RNN, LSTM, GRU, word embedding, and the attention mechanism, which have played a crucial role in advancing the field of MT. Furthermore, we present the fundamental architecture used in this field. To wrap up the chapter, we explore some of the metrics that are commonly used to evaluate MT systems, such as BLEU and METEOR. Overall, the third chapter of our thesis aims to provide a comprehensive understanding of the various approaches to MT and the key technologies involved. By exploring the nuances and complexities of this field, we give a deeper appreciation for the ongoing advancements and breakthroughs that are driving the field of MT forward.

In the fourth chapter of our thesis, titled 'Arabic Dialect Neural Machine Translation: An Overview and Contribution', we focus on the specific challenges presented by Arabic dialects. In particular, we present an overview of the works that have been done in the context of Arabic dialect NMT, including Arabic dialect corpora. Finally, we present our contribution to the field of NMT, which focuses specifically on translating Algerian and Arabic dialects using NMT models. Our research aims to address the unique challenges presented by these dialects, such as the need for specialized training data and the incorporation of region-specific grammar and vocabulary. Overall, the fourth chapter of our thesis provides a comprehensive overview of the latest advancements in the field of NMT, with a particular focus on the challenges and opportunities presented by Arabic dialects.

In the fifth chapter, titled 'Transfer Learning in Neural Machine Translation: An Overview and Contribution', we delve into the application of transfer learning to NMT, with a focus on improving the quality of MT. We begin by providing an overview of transfer learning and its applications in the field of machine learning. We discuss the setting and approaches of transfer learning and how we can integrate these approaches into NMT. We highlight the potential benefits of using transfer learning in NMT, such as the ability to improve performance with limited training data and the ability to transfer knowledge learned from one language pair to another. Next, we present some of the works that were applied in the context of Arabic dialect transfer learning NMT. Finally, we present our contributions to the fields of transfer learning and NMT, specifically in the context of translating the Algerian dialect. Our research focuses on using the transfer learning approach to improve the quality of MT for this dialect, which poses unique challenges due to its distinct grammar, syntax, and vocabulary. Our approach utilizes transfer learning techniques to improve the accuracy and efficiency of NMT for the Algerian dialect. Overall, the fifth chapter of our thesis provides a comprehensive overview of the transfer learning approaches for NMT as well as our contributions to the field.

Our thesis ends in **Chapter 6 as a conclusion**, where a review of the whole NMT applied system and the proposed contribution is discussed. After that, some insights on practical applications were presented. Finally, some limitations and directions for future work were pointed out.

Figure 1.1 provides the map of our thesis, which is structured into sections that overlaps, where, at first, a study of the Algerian dialect is offered, which is the dialect to be translated. In addition, we continue discussing MT and its techniques, then we go into further depth on one of the MT techniques, which is NMT. Furthermore, we open the door to the problem of the inability to translate low-resource Algerian dialects, from this point, we integrate transfer learning with NMT and solve this problem.

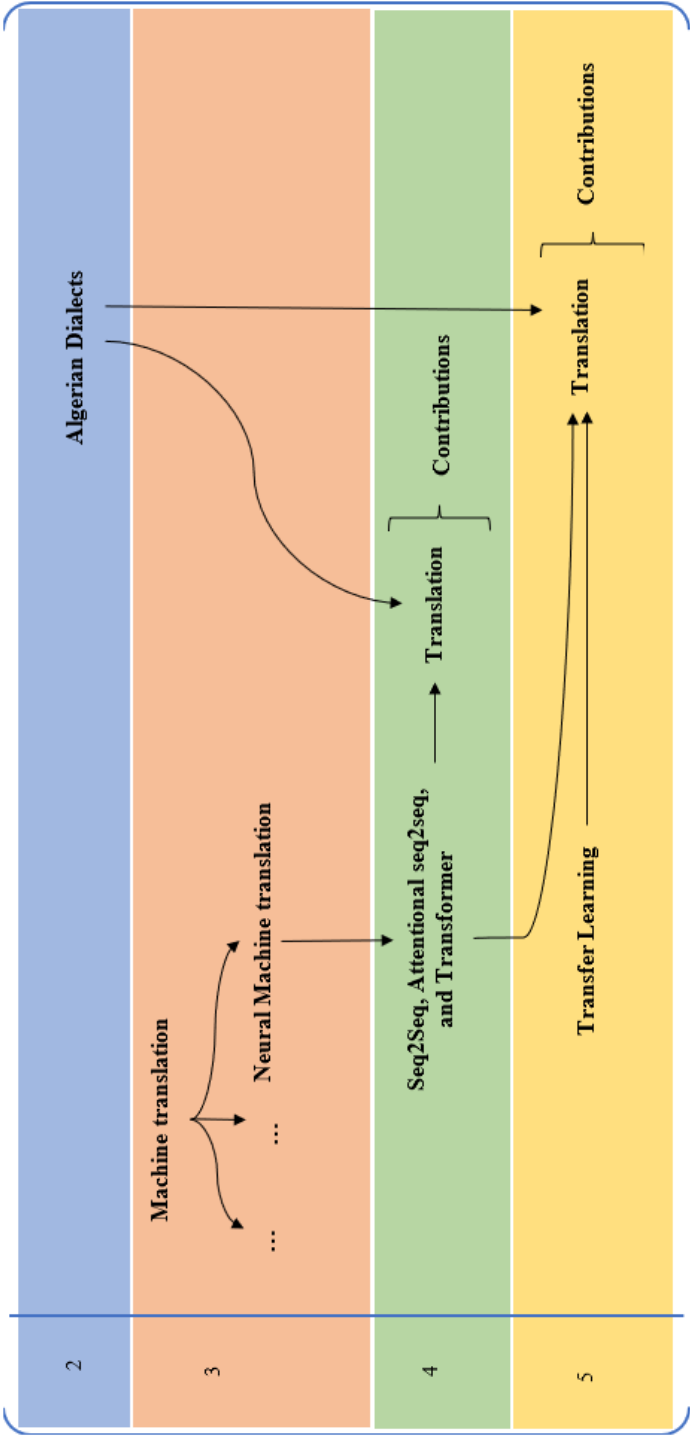


Figure 1.1: Thesis Map.

2

Algerian Dialect : A Linguistic Overview

2.1 Introduction

This chapter provides a comprehensive overview of the Algerian dialect, examining its linguistic features in lexicology, phonology, morphology, and syntax. The chapter is divided into five sections, each focused on a specific aspect of the dialect.

Section 2.2 explores the lexicology of the Algerian dialect, examining its vocabulary and the diverse linguistic influences that have shaped it over time.

Section 2.3 delves into the phonology of the dialect, highlighting its unique sound system and distinctive features.

Section 2.4 focuses on the morphology of the Algerian dialect, analyzing its complex system of word formation and inflectional processes, including the use of pronouns, verb conjugation, and declension in Algerian dialect. In this section, our focus lies on the ANB and ALG regional dialects from the Algerian dialect.

Section 2.5 presents the syntax of Algerian dialect, discusses the different types of sentences and constructions in the Algerian dialect, highlighting the unique features of its sentence structure. Finally, Section 2.6 provides a conclusion to the chapter.

2.2 Lexicology

Lexicology is the study of the vocabulary, words, and sentences used in a language. In the context of this section, our focus will be on the lexicology of the Algerian dialect,

examining its unique vocabulary, word formations, and sentence structures.

The goal of this section is to provide a thorough overview of the Algerian dialect's lexicon, emphasizing its complexity and diversity, and offering insights into its unique features and characteristics.

2.2.1 Vocabulary

In order to gain an understanding of the formation of the Algerian dialect vocabulary, a review of the region's history, culture, and politics is required. This review covers the impact of factors such as colonialism, migration, trade, and religion on the language's evolution. A study of the historical record will provide valuable insight into the dialect's vocabulary formation, and it is presented in this section.

The Tamazight, the original inhabitants of Algeria are divided into several tribes including Shawiya, Touareg, Beni Mzab, and Kabyle [12]. That's why, the Chaoui, Kabyle, Tergui, and Mozabite dialects are all part of the Tamazight language [13].

The spread of Arabic in Algeria started with Arab tribe settlements in the 7th century [14]. The influence of Islam and Arab political and economic control facilitated the Arabization of the region, leading to the widespread adoption of Arabic as the main language [15, 16]. Nevertheless, certain indigenous dialects, such as Kabyle and Chaoui, persist among descendants of the original Berber-speaking population [17, 18].

Throughout history, The Arabic Algerian dialect has undergone changes and has been influenced by different languages. Spanish words were introduced during Spanish colonial rule in the 16th century [18–20]. Turkish words were added during the Ottoman occupation [18, 19]. French words were adopted and adapted during French colonial rule [16, 20].

Table 2.1 showcases various loanwords from various languages that were incorporated into the Algerian dialect's vocabulary over time (the used words are taken from [18, 19]). These loanwords have enriched the linguistic landscape of the Algerian dialect, adding a diverse range of vocabulary to the language. This is a common phenomenon in languages, as they evolve and grow through exposure to and influence from other cultures and languages.

2.2.2 Semantics

Algerian dialects exhibit a variety of common linguistic features that can be observed in MSA, as well as other languages and dialects. These shared characteristics encompass the presence of polysemic words with multiple meanings, the utilization of synonyms denoting similar concepts, and the utilization of antonyms expressing opposite ideas. Within the Algerian dialect:

- **Polysemic:** we find numerous words that possess multiple meanings. For instance,

Table 2.1: Loanwords in Algerian Dialect Vocabulary

From	Loanword	English Translation
Spanish language	‘مزيرية’-	Misery
	‘ديوانة’-	Customs
	‘بطاطا’-	Potato
Turkish language	‘طبسي’-	A plate
	‘قهواجي’-	Cafe-boy, cafe-owner
	‘زردة’-	Feast
French language	‘تيكي’-	Ticket
	‘كازيرنة’-	(Military) Barracks
	‘بقیطة’-	Loaf of bread

the word ‘عين’ (ain). This term embodies a fascinating polysemy, as it can refer to a faucet used for controlling the flow of liquid, while also symbolizing the human eye as the organ of vision. The context in which ‘عين’ is used determines its intended meaning.

- **Synonyms:** The Algerian dialect employs synonyms to describe similar concepts. For instance, when referring to the concept of ‘rain,’ speakers can choose between two synonymous words: ‘النو’ (ennaw) and ‘لمطر’ (lmatar). Both terms convey the idea of precipitation, allowing individuals to express themselves using different linguistic choices.
- **Antonyms:** The Algerian dialect encompasses the usage of antonyms to express opposite meanings. For instance, consider the words ‘الليل’ (ellil) and ‘النهار’ (ennahar). ‘الليل’ signifies ‘night’, while ‘النهار’ refers to ‘daytime’. These two words stand as antonyms, conveying the contrasting concepts of darkness and daylight.

2.2.3 Lexical Categories

In linguistics, lexical categories refer to the different classes of words that make up a language. These categories provide a way to organize and classify words based on their meaning and function within a sentence. The lexical categories in the Algerian dialect include nouns, verbs, pronouns, and particles [21].

Nouns refer to people, places, things, or ideas and can be proper or common, singular or plural, and used as the subject or object of a sentence. Verbs refer to actions or states of

being and can be conjugated to reflect different tenses. Pronouns are words used to replace nouns and include four types: personal, possessive, adjective, and demonstrative pronouns.

Particles are used to express relationships between elements in a sentence and include prepositions, conjunctions, and quantifiers. Prepositions indicate the relationship between a noun or pronoun and other elements in a sentence, such as 'في' (in), 'عند' (at), and 'بين' (between). Conjunctions connect words, phrases, or clauses, such as 'و' (and), 'أو' (or), and 'والمبعد' (after). Quantifiers indicate the amount or quantity of something, such as 'قاع' (some), 'شوية' (a few), and 'كلش' (all). Interrogative used to form a question, such as 'وعلاه' (why), 'شحال' (how much), and 'شكون' (who). In the context of the Algerian dialect, as well as MSA, particles can also refer to words that show grammatical relationships between other words in a sentence but do not have an independent meaning. They are used to indicate direction, ask questions, or connect words or phrases, such as:

- The particle 'إلى' is employed to indicate direction or movement towards a specific destination. For instance, a phrase like 'روح إلى السوق' means 'Go to the market'.
- The particle 'وعلاه' is used in the Algerian dialect to form questions. For instance, the question 'وعلاه روجتي؟' means 'Why did you leave?'.
- The particle 'و' is utilized to connect words or phrases. For instance, 'جيب الخبز واللحم' means 'Bring bread and meat'.

2.3 Phonology

Phonology is the study of the sound systems of languages. Algerian dialect, like MSA, has a rich and diverse sound system. This section is divided into three main sub-sections: the phonetic inventory, phonological processes, and word pronunciation. These sub-sections will provide an in-depth analysis of the sounds used in the language and how they shape the pronunciation of words in Algerian dialects.

2.3.1 Phonetic Inventory

The phonetics inventory in the Algerian dialect involves examining the sounds and manner of pronunciation. It encompasses consonants, long and short vowels, double case endings (tanween), and syllabification marks, which are all present in MSA. However, the Algerian dialect differs from MSA in that it has discarded the double-case endings while retaining the remaining phonetic elements. Consequently, the phonetic inventory of this dialect distinguishes it from MSA.

Table 2.2: Phonetic table of Algerian Dialect [22].

Letter	Phonome	Letter	Phonome
Consonants			
أ	ʔ	ص	ʃ
ب	b	ض	ɖ
	p		
ت	t	ق	q
	t _s		g
	tâ		k
			ʔ
ث	θ	ط	ʈ
	t		
ج	ǧ	ظ	ɟ
	ʒ		ɖ
ح	ħ	ع	ʕ
خ	x	غ	ɣ
			q
د	d	ف	f
			v
ذ	ð	ك	k
	d		
ر	r	ل	l
ز	z	م	m
س	s	ن	n
ش	ʃ	ه	h
و	w	ي	j
Vowels			
Short		Long	
Fatha	a	ا	aa
Damma	u	و	uu
Kasera	i	ي	ii
Syllabification marks			
Shedda	Double letter	Sukun	Glotal stop

The long vowels used in the Algerian dialect include ‘ا’, ‘و’ and ‘ي’, which are pronounced as ‘با’ /baa/, ‘بو’ /buu/ and ‘بي’ /bii/ when added with the letter ‘ب’ /b/. These vowels are pronounced for a longer duration and are used to distinguish different words with similar pronunciation.

The short vowels in the Algerian dialect are indicated by diacritical marks placed above or below the Arabic letters, and they serve to clarify the pronunciation of the letters. These short vowels include Fatha /a/, Damma /u/, and Kasra /i/ and are pronounced with ‘ب’ /b/ as follows: ‘بَ’ with fatha is pronounced /ba/, ‘بُ’ with damma is pronounced /bu/.

Syllabification marks in the Algerian dialect include "shedda" and "sukun". "Shedda" doubles a consonant, such as ‘ب’ with shedda pronounced /bb/ and is usually combined with a vowel diacritic, such as ‘بَ’ pronounced /bba/ [23]. "Sukun" indicates the absence of a vowel, such as in ‘بْ’ pronounced /b/, and reflects a glottal stop [18, 23]. These syllabification marks are used to help clarify the pronunciation of words and to distinguish different words with similar pronunciation.

Additionally, Algerian dialect also includes phonemes from other languages, such as /g/, /p/ and /v/. These additional phonemes are used mainly in words borrowed from French, such as /g/ in the word ‘قاع’ (means ‘All’), the phoneme /p/ as in word ‘بومبة’ (means ‘pump’) and /p/ as in word ‘فليضة’ (means ‘bag’) [24]. Table 2.2 presents an inclusive compilation of the sounds and pronunciations found in the Algerian dialect. It provides a comprehensive understanding of the sounds and pronunciation used in this dialect, includes all the phonemes of the MSA, and also includes additional phonemes from other languages, making the Algerian dialect unique compared to MSA.

2.3.2 Phonological Process

The phonological processes refer to the ways in which sounds are produced, modified, or manipulated in a language. These processes play a crucial role in how words are perceived and understood by listeners, and they can vary greatly from one language or dialect to another. In the Algerian dialects, two main processes can be observed in phonology: assimilation and reduction.

Assimilation is the process of adjusting sounds to make them more similar to the sounds that are present in the surrounding environment. For instance, the pronunciation of the letter ‘ت’ in the Constantine region of Algeria is /t_s/ while in the Skikda region, it is pronounced as /t_h/. The phoneme ‘ث’ is pronounced as /θ/ in the Annaba and Guelma regions. Additionally, /ʒ/ in the Algeria and Oran regions [18, 22].

Reduction, on the other hand, refers to the simplification of sounds, making them shorter or less distinct. This process can be observed in the pronunciation of certain phonemes that are hardly used such as ‘ظ’ and ‘ذ’. In many dialects, these phonemes

Table 2.3: Differences in word pronunciation exist between the Algerian dialect.

case	Algerian Dialect	Pronunciation	MSA	Pronunciation	English Translation
1. Verb	شرب	šaraba	شرب	šaraba	Drink
Noun	مكتبة	maktaba	مكتبة	maktaba	Library
2. Verb	بقي	bqa	بقي	baqija	Stay
Noun	شمس	lqa	شمس	šams	Sun
3. Verb	لقي	lqa	وجد	wağada	Found
Noun	قهوة	qahwa	مقهى	maqha	café
4. Verb	شارجا	šarğa	شحن	šahn	Charge
Noun	بسكيات	bisqiilat	دراجة	daraja	Bicycle

are pronounced instead of other phonemes, for example, ‘ظ’ is pronounced as /d/, ‘ذ’ is pronounced as /d/, and ‘ث’ is pronounced as /t/. The phoneme ‘غ’ is pronounced as /q/ in the Aguatte dialect [18, 22].

2.3.3 Pronunciation

In terms of pronunciation, the Algerian dialect has some differences compared to MSA. One of the most noticeable differences is the pronunciation of certain letters or sounds as presented in section 2.3.1.

When it comes to differences in verbs and nouns between the Algerian dialect and MSA, there are four main cases [18]. The first case involves identical letters with the same diacritical marks, as seen in many Arabic dialects. The second case includes identical letters with differing diacritic marks, which can lead to variations in the pronunciation of vowels. The third case involves letter modifications or omissions, which can result in the deletion or addition of certain sounds or letters in the Algerian dialect. Lastly, the use of loanwords from other languages can also contribute to differences in vocabulary and pronunciation between the Algerian dialect and MSA.

A comparison of these differences in verbs and nouns between the Algerian dialect and MSA can be found in Table 2.3.

2.4 Morphology

In this section, we will explore the morphology of the Algerian dialect. We will delve into its various aspects such as word formation, pronouns, verb conjugation, and word declension. These linguistic features help us to understand the structure and form of words in the Algerian dialect and how they are used in speech and writing.

2.4.1 Word Formation

Word formation is an integral aspect of language, and refers to the process of creating new words within a language, either to meet the changing needs of its speakers or to reflect new experiences and concepts. The Algerian dialect, like other languages, also undergoes word formation processes to reflect the changing needs and realities of its speakers.

This process can occur through a variety of linguistic mechanisms, including inflection, derivation, compounding, clipping, and the use of acronyms and abbreviations. The inflection involves adding affixes to nouns and verbs (root) to indicate various grammatical features, such as number, case, gender, tense, mood, etc. For instance, adding the prefix 'ي' to the verb 'كتب' (means 'write') makes it "يكتب", indicating present tense, and adding the suffix 'و' makes it 'كتبو', indicating past tense. The noun 'معلم' (teacher) becomes 'معلمين' (teachers) in its plural form.

While, the derivation can be formed by the addition of affixes, such as prefixes, suffixes, and infixes, to root words. In the Algerian dialect, for instance, the verb 'كتب' can be transformed into various nouns through the addition of different affixes. For example, adding the infix 'ل' results in the noun 'كتاب' (book), while adding the prefix 'م' produces 'مكتب' (desktop). The noun 'مكتبة' (library) is formed by adding both the prefix 'م' and the suffix 'ة'. Additionally, the noun 'كتيبة' (writing) can be formed by adding the infix 'ي' and the suffix 'ة'.

Furthermore, in compounding and clipping new words are created by combining existing words or by shortening them. Compounding combines two or more words to create a new word with a different meaning, while clipping shortens a word to create a new word [25]. Examples of Compounding are 'بو جمعة (جمعة + بو)' (The word 'بو' signifies 'Friday,' 'جمعة' signifies 'father,' and 'بو جمعة' denotes a person's name) and examples of Clipping are 'صبح' (shortened form of 'صباح' (morning)).

2.4.2 Pronouns

In Algerian dialect like in MSA, the pronouns play an important role in expressing grammatical relationships and avoiding repetition. They can refer to people, places, things, or ideas, and there are several types of pronouns in the Algerian dialect like any language, including personal, possessive, adjective and demonstrative pronouns.

2.4.2.1 Personal Pronouns

Personal pronouns are used to refer to the speaker, the person being addressed, and others. The forms of personal pronouns change depending on the person, number, gender, and case.

Table 2.4: Personal pronouns in Algerian Dialect and MSA.

		Algerian Dialect	MSA	Eng
First Person	Singular	أنا	أنا	I
	Plural	حنا	نحن	We
Second Person	Singular Masculine	انتَ	انتَ	You
	Singular Feminine	انتِ	انتِ	
	Dual	أنتوم / أتم	انتما	
	Plural Masculine		أنتم	
	Plural Feminine		أنتن	
Third person	Singular Masculine	هو	هو	He
	Singular Feminine	هي	هي	She
	Dual	هم / هوم	هما	They
	Plural Masculine		هم	
	Plural Feminine		هن	

The personal pronouns used in the Algerian Dialect are different from those used in MSA (as shown in Table 2.4). The first-person singular pronouns are the same in both MSA and Algerian dialect, while the first-person plural is different ('نحن' in MSA and 'حنا' in Algerian dialect). MSA has dual pronouns for the second and third persons ('انتما' and 'هما'), which are not present in Algerian dialect. For the second and third persons, the singular pronouns are the same in both MSA and Algerian dialect, but the plural pronouns are different. MSA has separate plural pronouns for masculine ('أنتم') and feminine ('أنتن') second person, while Algerian dialect has a combined pronoun ('أتم') for both genders. Similarly, MSA has separate plural pronouns for masculine ('هم') and feminine ('هن') third person, while Algerian dialect has a combined pronoun ('هوم / هم') for both genders.

2.4.2.2 Possessive and Adjective Pronouns

Possessive pronouns and adjective pronouns help to express ownership and possession. Possessive pronouns replace a noun that shows possession. These pronouns demonstrate ownership and avoid the repetition of nouns. On the other hand, adjective pronouns, also known as possessive adjectives, modify a noun and come before it to show possession. These pronouns provide additional information about the noun being referred to, making it clear who or what it belongs to.

The Algerian dialect shares similarities with MSA in terms of the use of possessive

Table 2.5: Possessive pronouns in Algerian Dialect and MSA.

		Algerian Dialect	MSA	Eng
First Person	Singular	تاعي	لي	Mine
	Plural	تاعنا	لنا	Ours
Second Person	Singular Masculine	تاعك	لكَ	Yours
	Singular Feminine		لكِ	
	Dual	تاعكم	لكما	
	Plural Masculine		لكم	
	Plural Feminine		لكن	
Third Person	Singular Masculine	تاعو	له	His
	Singular Feminine	تاعها	لها	Hers
	Dual	تاعهم	لهما	Thiers
	Plural Masculine		لهم	
	Plural Feminine		لهن	

pronouns and adjectives as shown in Tables 2.5 and 2.6. However, the forms used in Algerian dialect are distinct from those used in MSA. In Algerian dialect, possessive pronouns have a different writing form compared to MSA. This can be seen in the comparison Table 2.5 of the forms used in the two languages.

It's also worth mentioning that in Algerian dialect, the singular masculine and feminine second-person possessive pronoun is expressed using the same form 'تاعك'. This is in contrast to MSA where the singular masculine second-person possessive pronoun is pronounced "laka" and the singular feminine second-person possessive pronoun is pronounced "laki," despite having the same written form.

Nevertheless, In the Algerian dialect, possessive adjectives are the same as in MSA, as shown in table 2.6. However, there are no possessive pronouns for second and third-person dual and plural masculine and feminine forms such as for Possessive Pronouns.

2.4.2.3 Demonstrative Pronouns

Demonstrative pronouns are a type of pronoun used to point out or refer to specific things or people. In a sentence, demonstrative pronouns take the place of a noun and demonstrate its proximity in time or space to the speaker. Demonstrative pronouns play an important role in communicating specific details about the objects or individuals being referred to and can help to create a clear and accurate picture for the listener or reader.

The use of demonstrative pronouns differs between Algerian Dialect and MSA. The difference can be summarized as follows:

Table 2.6: Possessive adjective in Algerian Dialect and MSA.

		Algerian Dialect	MSA	Eng
First Person	Singular	كتاي	كتاي	Mine
	Plural	كتابنا	كتابنا	Ours
Second Person	Singular Masculine	كتابك	كتابك	Yours
	Singular Feminine		كتابك	
	Dual	كتابكم	كتابكما	
	Plural Masculine		كتابكم	
	Plural Feminine		كتابكن	
Third Person	Singular Masculine	كتابو	كتابه	His
	Singular Feminine	كتابها	كتابها	Hers
	Dual	كتابهم	كتابهما	Thiers
	Plural Masculine		كتابهم	
	Plural Feminine		كتابهن	

Table 2.7: Proximal and Distal Demonstrative Pronouns in Algerian Dialect and MSA.

			Algerian Dialect	MSA	Eng
Proximal	Singular	Masculine	هادا	هذا	this
		Feminine	هادي	هذه	this
	Dual	Masculine	هادو	هاذان	these
		Feminine		هاتان	
	Plural			هؤلاء	
Distal	Singular	Masculine	هداك	ذلك	that
		Feminine	هاديك	تلك	that
	Dual	Masculine	هادوك	ذانك	those
		Feminine		تانك	
	Plural			اولئك	

- In Algerian dialect, each singular masculine or feminine proximal or distal form has its own demonstrative pronoun, which is different in writing and pronunciation from MSA.
- In Algerian dialect, the same pronoun هادو is used for dual (masculine and feminine) and plural, whereas in MSA, two different pronouns are used for dual masculine and feminine forms: هادان for dual masculine, and هاتان for dual feminine. The same applies to distal pronouns, with هادوك used for both masculine and feminine distal forms as shown in Table 2.7.

2.4.3 Verb Conjugation

The Algerian dialect has its own set of rules for conjugating verbs based on different aspects such as voice (active or passive), tense (past, present, or future), a person (feminine or masculine), and number (singular or plural). This conjugation is somewhat similar to the conjugation in MSA but with its own unique variations.

In this section, we will explore the verb conjugation in the past, present, future, and imperative moods in the Algerian dialect, with a focus on two specific dialects: the ALG dialect and the ANB dialect. The differences in pronoun usage between the Algerian dialect and MSA will be reflected in the conjugation patterns.

For clarity, the verb أكل (meaning "ate") will be utilized as a constant example throughout our comparison of verb conjugation between the Algerian dialect and MSA. The verb أكل has an identical written form in both the Algerian dialect and MSA, making it a suitable choice to emphasize the differences in conjugation. This will enable us to focus specifically on the variations in conjugation patterns between the two forms of Arabic.

2.4.3.1 Past Tense

Verb conjugation in the past tense plays a vital role in Algerian dialect inflection and provides a way for speakers to indicate that an action has already happened and is finished. This tense is achieved by making modifications to the root form of the verb through the addition of different prefixes and suffixes.

The conjugation of verbs in the past tense in the Algerian dialect is distinct from the conjugation in MSA. In the following, we will provide the key differences in verb conjugation in the past tense in the Algerian dialect and MSA (Table 2.8 shows the conjugation of the verb أكل in the past in Algerian dialect and MSA).

- Writing form: The writing form of verbs in the past tense in the Algerian dialect is different from that in MSA. For example, in the Algerian dialect, the letter أ (alif) is often removed from the root form of the verb, such as in the verb أكل (ate).

Table 2.8: Past tense Verb conjugation in Algerian Dialect and MSA.

		Algerian Dialect		MSA	Eng
		ALG	ANB		
First Person	Singular	أنا كليث	أنا كليث	أنا أكلت	I ate
	Plural	حنا كلينا	حنا كلينا	نحن أكلنا	we ate
Second Person	Singular Masculine	انت كليث	انت كليث	انت أكلت	you ate
	Singular Feminine	انت كليث	انت كليث	أنت أكلت	
	Dual	أنتوم كليتو	أنتوم كليتو	انتما أكلتما	
	Plural Masculine			أنتم أكلتم	
	Plural Feminine			أنتن أكلتن	
Third Person	Singular Masculine	هو كل	هو كل	هو أكل	he ate
	Singular Feminine	هي كلث	هي كلث	هي أكلت	she ate
	Dual	هوم كلاؤ	هوم كلاؤ	هما أكلا	they ate
	Plural Masculine			هم أكلو	
	Plural Feminine			هن أكلن	

Table 2.9: Present tense verb conjugation in Algerian Dialect and MSA.

		Algerian Dialect		MSA	Eng
		ALG	ANB		
First Person	Singular	نوكل	نوكل	أنا أكل	I eat
	Plural	نوكلو	نوكلو	نحن نأكل	we eat
Second Person	Singular Masculine	توكل	توكل	انت تأكل	you eat
	Singular Feminine	توكل	توكل	أنت تأكلين	
	Dual	توكلو	توكلو	انتما تأكلان	
	Plural Masculine			أنتم تأكلون	
	Plural Feminine			أنتن تأكلن	
Third Person	Singular Masculine	يوكل	يوكل	هو يأكل	he eats
	Singular Feminine	توكل	توكل	هي تأكل	she eats
	Dual	يوكلو	يوكلو	هما يأكلان	they eat
	Plural masculine			هم يأكلون	
	Plural Feminine			هن يأكلن	

- First-person conjugation: To form the past tense in the Algerian dialect, various prefixes are added to the root form of the verb. For singular, the prefix **آيت** is added, and for plural, the prefix **ينا** is added.
- Second-person conjugation: The conjugation of verbs in the past tense in the Algerian dialect is different between the ALG dialect and the ANB dialect. In the ALG dialect, the prefix **آيت** is added for the singular masculine, and the prefix **يت** is added for the singular feminine. In the ANB dialect, the same prefix **يت** is added for both singular masculine and feminine, and the prefix **يتو** is added for dual and plural (masculine and feminine).
- Third-person conjugation: To form the past tense in the third person in the Algerian dialect, various prefixes are added to the root form of the verb. The prefix **أ** is added for singular masculine, the prefix **ات** is added for singular feminine, and the prefix **او** is added for dual and plural (masculine and feminine).

2.4.3.2 Present Tense

The conjugation of verbs in the present tense in the Algerian dialect differs from that in MSA. In MSA, the present tense is formed by adding the prefixes **أ**, **ن**, **ي** and **ت** to the verb. However, in Algerian dialect, only the prefixes **ن**, **ي** and **ت** are used to create the present tense form of the verb.

In addition to the prefixes, there are also suffixes added to express the gender (feminine or masculine) and number (singular or plural) of the verb. These suffixes are as follows (as shown in Table 2.9):

1. Singular Feminine Suffix:

- In Algerian dialect, the diacritic mark 'kesra' is added at the end of the verb to express singular feminine in the second person. For instance, the verb **أكل**, will be **توكلي** in its feminine form.
- In MSA, the suffix **ين** is added to the verb to express singular feminine in the second person. For instance, the verb **أكل** will be **تأكلين** in its feminine form.

2. Dual and Plural Suffixes:

- In Algerian dialect, the suffix **و** is added to the verb to express dual and plural in the first, second, and third person. For instance, the verb **أكل** will be **توكلو** in its plural forms.

Table 2.10: Future tense verb conjugation in Algerian Dialect and MSA.

		Algerian Dialect		MSA	Eng
		ALG	ANB		
First Person	Singular	درك نوكل	درك نوكل	انا سأكل	I will eat
	Plural	درك نوكلو	نوكلو	نحن سنأكل	we eat
Second Person	Singular Masculine	درك توكل	توكل	انت تأكل	you will eat
	Singular Feminine	درك توكلي	درك توكلي	أنت ستأكلين	
	Dual	درك توكلو	درك توكلو	انتما ستأكلان	
	Plural Masculine			أنتم ستأكلون	
	Plural Feminine			أنتن ستأكلن	
Third Person	Singular Masculine	درك يوكل	درك يوكل	هو سيأكل	he will eat
	Singular Feminine	درك توكل	درك توكل	هي ستأكل	she eats
	Dual	درك يوكلو	درك يوكلو	هما سيأكلان	they will eat
	Plural masculine			هم سيأكلون	
	Plural Feminine			هن سيأكلن	

- In MSA, the suffixes **ان**, **ون** and **ن** are added to the verb to express dual and plural in the second and third person. For instance, **تأكلان**, **تأكلون**, and **تأكلن** are dual and masculine and feminine plural of the verb **أكل**, respectively.

2.4.3.3 Future Tense

To express the future tense in MSA, there are two ways: the near future and the far future. For the near future, the prefix **س** is added to the verb in its present tense form. In both ALG and ANB, the word **درك** (meaning 'now') or the letter **ح** is employed before the present tense conjugated verb as an ante-position to convey the concept of the near future. For instance, **سأكل** and **درك نوكل** are the first-person future form of the verb **أكل** in MSA and Algerian dialect, respectively.

For the far future, MSA uses the word **سوف** (meaning 'will') as an ante-position before the verb conjugated in the present tense. In contrast, the Algerian dialect has different methods of expressing the far future. In the ALG dialect, the word **اومبعد** (meaning 'later') is added as an ante-position before the verb conjugated in the present tense. In the ANB dialect, the word **مبعد** (meaning 'later') is added as an ante-position before the verb conjugated in the present tense. For instance, the verb **سوف أكل** will be **أكل** in first person

Table 2.11: Imperative Mood verb conjugation in Algerian Dialect and MSA.

		Algerian Dialect	MSA	Eng
Second Person	Singular Masculine	أخرج	أخرج	Go out
	Singular Feminine	أخرجي	أخرجي	
	Dual	أخرجو	أخرجا	
	Plural Masculine		أخرجوا	
	Plural Feminine		أخرجن	

future form in MSA. While, it will be *أومبعد نوكل* and *مبعد نوكل* in is first person future form in ALG dialect and ANB dialect, respectively.

It is worth noting that while the Algerian dialect has distinct forms of verb conjugation in the present tense compared to MSA, the conjugation rules for the future tense are similar in both ANB and ALG dialects. Table 2.10 presenting better illustration of the verb *أكل* (meaning ‘ate’) conjugated in the future tense in both ALG and ANB dialects.

2.4.3.4 Imperative Mood

The imperative mood is used to give a command or instruction. In MSA, the conjugation of verbs in the imperative mood is simple compared to other forms. It is achieved by using the root form of the verb with minimal modification. Table 2.11 shows the conjugation of the verb *خرج* (to go out) in the imperative mood, demonstrating the simplicity of the conjugation process in this mood.

The imperative mood in Algerian dialect, for singular masculine and feminine similar to that in MSA, the verb is used as is, with no changes to the spelling. However, in some cases, the pronunciation may differ. For example, the verb *رجع* (to go back) has different imperative forms in MSA and Algerian dialect. In MSA, the imperative form is *ارجع* and it is pronounced as /ʔirǧaʕ/. However, in the Algerian dialect, it is pronounced as /ʔarǧaʕ/.

On other hand, for a dual and plural masculine and feminine, the suffixes *ا*, *و*, and *ن* are added to the verb in MSA. While in Algerian dialect, the suffix *و* is add. For instance, *أخرجو* is the imperative mood of the verb *خرج* for dual and plural masculine and feminine.

2.4.4 Word Declension

Word declension is the process by which the form of a word changes to express grammatical relationships, such as number, gender, and case. In the context of the Algerian dialect, noun declension would be the primary focus of word declension.

Figure 2.1 provides a comprehensive explanation of the differences and similarities between the Algerian Dialect and the MSA in terms of their plural formation. Both Algerian

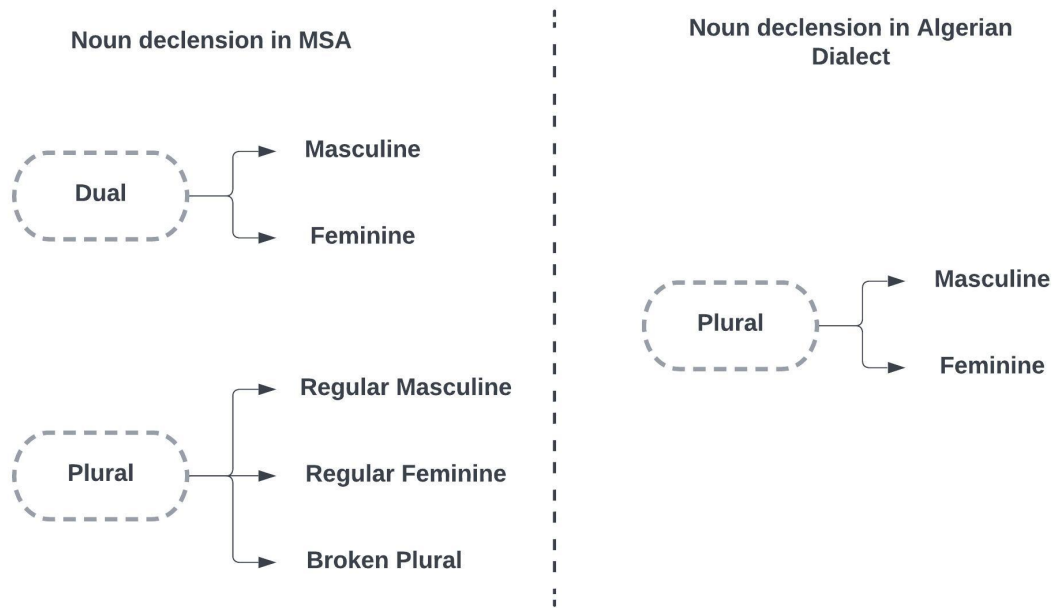


Figure 2.1: Difference between Noun Declension in Algerian Dialect and MSA.

dialect and MSA share the plural, but MSA has an added feature of the dual, which is not present in the dialect.

The dual is a linguistic category that is used to refer to two objects or persons. For instance, the word **ولد** means ‘boy’, and in dual form, it becomes **ولدان** which means ‘two boys’. In the Algerian dialect, it is customary to add the word ‘زوج’ as an ante-position to the plural noun to indicate the dual form.

The plural, on the other hand, is used to refer to more than three objects or persons. The plural in MSA is further divided into three categories, namely the regular masculine plural, the regular feminine plural, and the broken plural.

The regular masculine plural is formed without modifying the word, and by adding the suffix **ون** when the last letter of the word is discretized by damma (Nominative word). For example, the word **معلم** which means ‘teacher’ in Arabic become **معلمون** in its regular masculine plural form. In the case where the last letter of the word is discretized by fatha (accusative word), or kasra (genitive word) [26], the suffix **ين** is added, resulting in **معلمين**. However, in Algerian dialect, these rules are not followed, and the regular masculine plural is formed by adding the suffix **ين** without modifying the word. However, in Algerian dialect, these rules are not followed, and the regular masculine plural is formed by adding the suffix **ين** without modifying the word, for instance **معلم** become **معلمين** [18].

Regular feminine plural in algerian dialect and MSA, is obtained by adding the suffix **ات** without modifying the word. But if the end of the word is **ة**, the **ة** is omitted and replaced by **ات**. For instance, **معلمة** become **ملمات** [18].

The broken plural, as the name suggests, refers to a plural form that differs from its

singular form, either by excess or omission. This type of plural formation is unique to MSA and provides a more nuanced way of expressing the plural in Arabic [24].

2.5 Syntax

The syntax of the Algerian dialect section is concerned with the way words are organized to form phrases, clauses, and sentences. This includes analyzing word order, which refers to the arrangement of words within a sentence and the relationships they express. It also covers clause structure, which describes how clauses are formed by combining words and phrases to express a complete thought. Finally, sentence type refers to the different types of sentences based on their structure and function, such as declarative, interrogative, exclamatory, and imperative sentences.

2.5.1 Word Order

In Algerian Arabic dialects, the word order can be dynamic and adjust to the situation, with an emphasis on context and purpose. This allows for greater flexibility and precision in communication. The most common word orders used are Subject-Verb-Object (SVO) and Subject-Object-Verb (SOV), but the order can also be influenced by various factors such as the local language and culture, the speaker's education, social status, and the context of the language use. Some of the most common word orders in the Algerian Dialect include:

1. Subject-Verb-Object (SVO): This is the most basic word order in the Algerian dialect, where the subject of the sentence comes first, followed by the verb, and finally the object, such as in ' *طفل راح للمدرسة* ' (The child was going to school).
2. Subject-Object-Verb (SOV): In this word order, the subject comes first, followed by the object, and finally the verb. This word order is often used to place emphasis on the object, such as in ' *طفل للمدرسة راح* ' (The child was going to school).
3. Verb-Subject-Object (VSO): In this word order, the verb comes first, followed by the subject, and finally the object. This word order is often used in questions or in sentences where the focus is on the action being performed, such as in ' *راح طفل للمدرسة* ' (The child was going to school).
4. Object-Subject-Verb (OSV): In this word order, the object comes first, followed by the subject, and finally the verb. This word order is often used to place emphasis on the object, such as in ' *للمدرسة طفل راح* ' (The child was going to school).

5. Verb-Object-Subject (VOS): In this word order, the verb comes first, followed by the object, and finally the subject. This word order is often used in questions or in sentences where the focus is on the action being performed, such as in 'راح للمدرسة طفل' (The child was going to school).

Compared to Standard Arabic, the word order in Algerian Arabic is more flexible and adaptable.

2.5.2 Clause Structure

Algerian dialect is a flexible dialect share with MSA the structure for forming sentences. The clause structure in Algerian dialect can be divided into three categories: simple, compound, and complex.

A simple clause is the most basic form of a sentence and consists of a single verb and its subject. For example, "I eat" (أنا نوكل) this type of clause is used to make a simple statement or ask a simple question.

A compound clause is a more complex form of a sentence and consists of two or more simple clauses connected by a coordinating conjunction such as "and" (و) or "but" (لكن). For example, "I eat an apple and he drinks coffee." (أنا نوكل تفاحة و هو يشرب قهوة) Compound clauses are used to express more complex relationships between events or ideas, and can be used to make complex statements or ask complex questions.

A complex clause is a clause that has a main clause and one or more dependent clauses. Dependent clauses are clauses that cannot stand alone as a sentence and must be attached to a main clause to make sense.

For example, "When I finish my work, I will go home" (كيما نكمل خدمتي، نروح لدار) In this sentence, "when I finish my work" is a dependent clause, and "I will go home" is the main clause. Complex clauses are used to express relationships of time, condition, purpose, and cause between events or ideas.

It is worth noting that the flexibility of the Algerian dialect allows for a flexible word order, which allows speakers to express different meanings by rearranging the words in a sentence.

2.5.3 Sentence Type

Sentence type refers to the grammatical category a sentence belongs to, based on its structure and the type of information it conveys. Knowledge of the various sentence types, including Declarative, Interrogative, Imperative, Exclamatory, and Negative, is an essential aspect of effective communication in any language. Furthermore, the flexibility of the Algerian dialect enables the formation of sentences in any of these various sentence types.

Table 2.12: Interrogative Particles used in MSA, ALG and ANB Algerian Dialects.

MSA	ALG	ANB	Eng
‘متى’	‘وقتاش’	‘وكتاه’	When
‘ماذا’	‘واش’	‘واش’	What
‘لماذا’	‘وعلاه’	‘علاه’	Why
‘كيف’	‘كيفاش’	‘كيفاه’	How
‘أنى’	‘منين’	‘منين’	How
‘من’	‘شكون’	‘شكون’	Who
‘اين’	‘وين’	‘وين’	Where
‘كم’	‘شحال’	‘بقده’	How much
‘لمن’	‘لشكون’	‘تع شكون’ or ‘لشكون’	To whom

The declarative sentence expresses a statement and conveys information or makes a claim. It is indicated by ending with a period. Examples "السماء صافية" (The sky is clear) or "عندي كلب" (I have a dog). The purpose of a declarative sentence is to provide information or make a statement.

The imperative sentence in the Algerian dialect is used to express a command or request. Its structure is simple, starting with the verb in the imperative mood, followed by the object or a sentence. Example "جيب معاك الخبز" (Bring the bread with you). This arrangement clearly conveys the message, making it a straightforward and effective way of communicating commands and requests.

The interrogative sentences are used to ask questions. The Algerian Dialect has its own set of interrogative particles for constructing these sentences, which differs from the particles used in MSA. This distinction can be seen in Table 2.12, which displays the various interrogative particles used in MSA, ALG dialect, and ANB dialect, emphasizing the unique nature of the interrogative particles utilized in the Algerian dialect :

- "وقتاش" or "وكتاه": "when?" It's used to ask about the time when something happened or will happen.
- "واش": "what?" It's used to ask about the identity or nature of something.
- "وعلاه" and "علاه": "why?" It's used to ask about the reason or cause for something.
- "كيفاش" and "كيفاه": "how?" It's used to ask about the method or manner in which something is done.
- "منين": "how?" It's used to ask about the method or manner in which something is done.

- 'شكون': "who?" It's used to ask about the person who does or is responsible for something.
- 'وين': "where?" It's used to ask about the location of something.
- 'شحال' and 'بقداه': "how much?" It's used to ask about the quantity of something.
- 'لشكون' and 'تعا شكون': "for whom?" It's used to ask about the person for whom something is intended or done.

The exclamatory sentence is a type of sentence that is used to express strong emotions such as surprise, excitement, or emphasis. These emotions are conveyed in the sentence through the use of specific linguistic features and techniques. In MSA, the exclamatory sentence is formed by adding the exclamatory particle "أ" to the beginning of the sentence. This particle serves to indicate the emotional tone of the sentence and to emphasize the emotion being expressed. On the other hand, in Algerian Dialect, exclamatory sentences are often shown through the use of facial expressions and gestures. This type of non-verbal communication plays a significant role in conveying emotions in Algerian Dialect and is often used in conjunction with words and phrases to enhance the emotional impact of a sentence. In addition to non-verbal cues, exclamatory sentences in Arabic Dialect are also formed through the use of specific adjectives. These adjectives are used to indicate the exclamation sentiment and to enhance the emotional tone of the sentence. For example, "روعة" (good) and is used to express amazement or excitement, while "مليح" (beautiful) and is used to express admiration. These adjectives help to convey the emotional tone of the sentence and to emphasize the strength of the emotion being expressed.

The formation of negation sentence in MSA and Algerian Dialect is similar, but the negation particles used in Algerian dialect are different from those used in MSA. The negation particle in MSA is "لا", It is used to negate verbs, nouns, and adjectives. While in Algerian dialect, there are two negation particles:

- The negation particle "ما + ش" is used to negate verbal sentences and is formed by adding the prefix "ما" and the suffix "ش" to the verb. For example, the declarative form "لعب طفل" (The child played) becomes "مالعبش طفل" (The child didn't play.) in the negative form.
- The negation particle "ماشي" is used to negate nominal sentences. It can be placed before or after the subject. For example, the declarative form "اطفل مريض" (The child is sick) becomes "ماشي مريض طفل" or "طفل ماشي مريض" (The child is not sick) in the negative form. If the sentence begins with a pronoun, the particle "ماشي" can also be used to negate a verbal sentence.

2.6 Conclusion

The Algerian dialect of Arabic is a complex and distinct variety that reflects the diverse cultural and linguistic landscape of Algeria. This chapter has explored its lexicology, phonology, morphology, and syntax, highlighting the unique features of each component. In the realm of lexicology, we explored the historical development of vocabulary in Algerian dialects, as well as their semantics and lexical categories. Within phonology, we examined the phonetic inventory, phonological processes, and word pronunciation specific to the Algerian dialect. Turning to morphology, we presented aspects such as word formation, pronouns (personal, possessive, adjective, and demonstrative pronouns), verb conjugation (including past, present, and future tenses, as well as the imperative mood), and noun declension in the Algerian dialect. In the domain of syntax, we delved into word order, clause structure, and sentence types. By analyzing these linguistic elements, we can gain a deeper understanding of the dialect and its role in Algerian society, considering its differences from MSA.

On the other hand, MT faces a significant challenge in translating the Algerian dialect because of its uniqueness and complexity. Despite this challenge, it is crucial to translate this dialect to promote intercultural communication and understanding between Algerian society and other cultures. Failure to do so can lead to limited opportunities for exchange, stifling the potential for mutual enrichment and cultural exchange.

3

Machine Translation

3.1 Introduction

MT has become an indispensable tool in our rapidly globalizing world, enabling people to communicate and share information across language barriers. In this chapter, we delve into the complexities of machine translation, starting with an overview of the concept itself in section 3.2.

Next, we discussed the methods used to achieve it. We will provide the two main approaches to MT: rule-based, and corpus-based, and discuss their underlying principles and the mechanisms by which they work in sections 3.3, and 3.4, respectively.

In section 3.5, we explore the various metrics used to evaluate the quality of machine translation. These metrics, including BLEU and METEOR, provide a way to measure the accuracy and effectiveness of a particular approach. And, we close the chapter with a conclusion (section 3.6).

In summary, this chapter provides a comprehensive look into the world of machine translation, by providing a deeper understanding of the different methods and metrics used in machine translation.

3.2 What's Machine Translation

The translation process, whether carried out by humans or machines, follows two fundamental steps. The first step involves decoding the meaning of the source language (SL), and

the second step entails re-encoding this meaning into the target language (TL). Behind its appearance of simplicity is a sophisticated cognitive activity, the translator must thoroughly understand the meaning of the source text by analyzing and interpreting its features, such as grammar, semantics, syntax, idioms, and more. This requires a comprehensive knowledge of the SL and the cultural background of its speakers [27, 28].

MT system involves the automatic translation of text or speech from one SL to TL using computer algorithms. The process of MT typically involves three key steps, the MT system begins by analyzing the SL input and creating an internal representation of its meaning. This representation is then transformed and adapted to the structure of the TL.

Machine translation (MT) encompasses a range of methods used to translate text from SL to TL. These methods can be broadly classified into two main categories: RBMT, and corpus-based machine translation (CBMT).

RBMT heavily relies on the expertise of human linguists and translators who create a set of specific rules governing the translation process. In contrast, CBMT utilize parallel corpora, which are collections of translation samples created by human experts, to automatically acquire knowledge about translation patterns.

3.3 Rule Based Machine Translation

RBMT is one of the earliest strategies in the field of MT, dating back to the 1954s [29]. This system consists of a set of rules, a lexicon (bilingual or multilingual), and software programs to implement the rules. The rules are based primarily on the analysis of the SL and the generation of the TL in terms of grammar structures, such as syntax, semantics, morphology, part-of-speech tagging, and orthographic features [30, 31].

RBMT systems follow various translation approaches, including Direct, Transfer, and Interlingua. These approaches differ in the depth of analysis of the SL and the extent to which they aim to reach a source and TL-independent representation of meaning. The differences between these approaches can be visualized using the Vauquois Triangle [32] (shown in Figure 3.1), which shows the levels of analysis involved.

3.3.1 Direct Machine Translation

Direct MT, also known as word-based or dictionary-based translation, is a simple and straightforward approach to MT. The direct MT approach uses a bilingual dictionary and a morphology analyzer to translate SL text into the TL [33–35]. The process of direct MT can be broken down into four stages:

1. Morphology analysis: In this stage, the basic forms of the SL words are determined by eliminating inflections and resolving ambiguities.

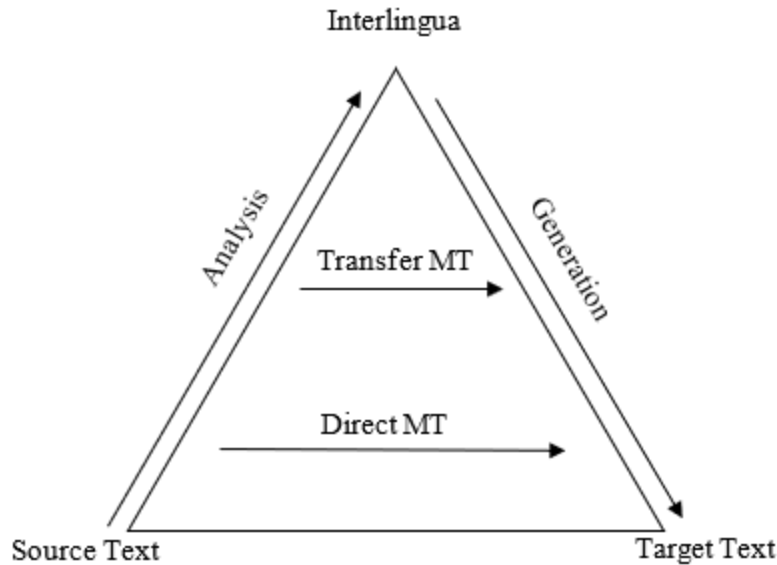


Figure 3.1: The Bernard Vauquois Pyramid of RBMT Approaches [32].

2. Bilingual dictionary lookup: The SL base forms are looked up in the bilingual dictionary, and comparable TL base forms are generated.
3. Rule application: Rules are used to make small grammatical changes to the TL word order and morphology generator.
4. Output: The final output is the translated text in the TL.

While direct MT may be adequate for simple, straightforward translations, it is limited in its ability to handle more complex linguistic structures and cannot produce high-quality translations in many cases. As a result, more advanced approaches, such as transfer-based and interlingua-based translation, have been developed to improve the quality and accuracy of MT.

3.3.2 Transfer Machine Translation

The Transfer MT Approach is a complex process that is influenced by the languages pair being used. This approach involves three distinct stages, each designed to address the structural differences between the SL and TL [36, 37].

- In the first stage, known as Analysis, the SL parser is utilized to create a syntactic representation of the SL sentence. This is an important step in ensuring that the text is structured in a way that can be effectively translated.
- The second stage, Transfer, involves converting the parsed text into equivalent representations that are optimized for the TL. This stage is critical in ensuring that the meaning and context of the original text are accurately conveyed in the TL.

- Finally, in the Generation stage, a TL morphological analyzer is used to generate the final text in the TL. This stage ensures that the output is grammatically correct and properly structured.

Despite its advantages, the Transfer Approach in MT also has its limitations. One of the main issues is the sheer number of rules that must be applied at each stage of the translation process. These rules cover everything from the source-to-target transfer, TL synthesis, and source-to-target analysis, making it challenging to keep the transfer modules as straightforward as possible. Additionally, the complexity of these rules can make it difficult to maintain consistency and accuracy in the translation output.

3.3.3 Interlingua Machine Translation

The Interlingua MT method is a more advanced and sophisticated approach to RBMT compared to Direct MT and Transfer MT. It involves converting the SL text into an intermediate representation known as an interlingua, which can then be translated into the TL. This intermediate stage, interlingua, is designed to be language-independent, making it possible to use the same interlingua for translation into multiple languages [38].

One of the major advantages of this system is that the interlingua becomes more valuable as the number of TLs increases. This makes it possible to achieve more efficient and accurate translations compared to other RBMT methods. However, despite its benefits, the Interlingua approach is not without its challenges. One of the main difficulties is extracting meaning from the SL text and converting it into the interlingua representation. This requires sophisticated linguistic models that can accurately analyze the SL text and convert it into interlingua. Another challenge of the Interlingua approach is semantic differentiation, which is specific to the TL. This makes it difficult to make the necessary distinctions between words and phrases in the TL, particularly when dealing with complex or abstract concepts. This can result in poor-quality translations that lack the nuance and context of the SL text [27].

3.4 Corpus Based Machine Translation

CBMT is a highly effective method in the field of MT that differs from the traditional rule-based approach by utilizing a parallel corpus, or a collection of texts in two or more languages, to generate translations.

CBMT systems are typically based on SMT, Example based machine translation (EBMT), or NMT techniques. The core idea behind CBMT is to leverage the patterns and statistical regularities observed in the bilingual or multilingual corpora to make translation decisions [39].

3.4.1 Statistical Machine Translation

SMT adopts a data-driven approach to the translation process. By harnessing extensive collections of parallel text data, this approach constructs statistical models with the ability to produce translations.

The SMT approach is based on the idea that translation can be modeled as a statistical reasoning problem [31, 40]. It utilizes parallel corpora of text in two different languages, with each sentence in the TL corresponding to a sentence in the SL, to derive statistical models that capture the patterns in the translation process. The resulting models are then used to generate translations of new text.

The SMT architecture is comprised of three key components: the language model, the translation model, and the decoder [31]. The language model calculates the probability of the TL $P(t)$, by modeling the probability distribution of words and sentence in the TL. The translation model, on the other hand, calculates the conditional probability of the TL output given the SL input $P(t|s)$, by modeling the probability distribution of TL words and phrases given their SL counterparts. Finally, the decoder uses the outputs of both models to determine the best possible translation $P(t|s)$, by maximizing the combined probability using the formula 3.1.

$$t = \arg \max(P(t|s) * P(t)) \quad (3.1)$$

In order for SMT to perform well, it requires a large amount of high-quality parallel text data. This is because the statistical models that form the foundation of the approach are derived from the analysis of this data. The more data that is available, the better the models will be, and the higher the quality of the translations will be. However, it is important to note that the approach can struggle with languages that have significantly different word orders, such as Japanese and European languages [27], and additional methods may need to be employed to improve the quality of the translations in these cases.

In essence, SMT is a powerful and flexible approach to MT compared to RBMT approaches, that has been successfully applied to a wide range of language pairs. Despite its limitations, it remains a popular and widely used method for generating high-quality translations.

3.4.2 Example Based Machine Translation

EBMT was pioneered by Nagao et al [41] and involves the use of a bilingual corpus with parallel texts as its primary source of knowledge. The approach relies on analogy-based translation, which leverages similar meanings and examples stored in a database [42, 43]. This database consists of parallel, aligned bilingual corpora where pairs of the SL and TL translations are stored. The alignments can occur at various levels of granularity, such

as sentence, or word level, among others. The EBMT process comprises three stages: matching, adaptation, and recombination [42, 44?].

The process of matching in EBMT is a key step that starts with breaking down the SL text into fragments based on the granularity of the system. The next step involves searching for examples from the database that closely match the SL fragments. Once the relevant examples have been selected, the corresponding TL fragments are extracted.

There are several methods for matching, including Partial Matching for Coverage, Structure-based Matching, Annotated Word-based Matching, Word-based Matching, and Character-based Matching [42]. The choice of method depends on the requirements of the EBMT system and may involve a combination of methods.

The next step in EBMT after matching is Adaption. If the matching process results in an exact match, the fragments are simply recombined to form the TL output. If the match is not exact, the corresponding TL portion of the relevant match is aligned with the specific portion of the SL.

The final step is Recombination, which involves combining the relevant TL fragments in a manner that results in a grammatically correct TL text.

EBMT is a highly appealing approach to translation due to its avoidance of the manual creation of rules. Despite its benefits, EBMT still faces some challenges, one of which is the need for analysis and generation modules to create the dependency trees required for both the examples database and for analyzing sentences. Another issue with EBMT is computational efficiency, especially when dealing with large databases, but this can be addressed through the use of parallel computation techniques [27].

3.4.3 Neural Machine Translation

The discovery of neural networks and advancements in deep learning has had a positive impact on various areas of Natural Language Processing (NLP), including MT. As a result, NMT [45–48] has become the dominant approach, replacing SMT as the mainstream method for language translation [49].

In real-time, NMT allows for the translation of millions of pieces of information with accuracy and reliability comparable to that of a human. The simplicity of NMT has contributed to its state-of-the-art performance across various language pairs [50], making it the key technology behind many commercial MT systems [51, 52]

In mathematical terms, the goal of NMT training is to estimate the probability p (Equation 3.2), of a target sentence y given a source sentence x . The training process involves providing the NMT model with a parallel corpus C , which consists of a series of sentence pairs (x, y) where x represents the input sentence, consisting of $x = \{x_1, \dots, x_n\}$, and y represents its corresponding translation, consisting of $y = \{y_1, \dots, y_m\}$. Both x and y are sequences of characters or words, with n being the length of the input sentence x and m

being the length of the output sentence y . It is worth noting that the lengths of x and y may not be the same.

$$p(y_1, y_2, \dots, y_m | x) = \prod_{t=1}^n p(y_t | y_{(<t)}, x) \quad (3.2)$$

The seq2seq model, first introduced in [47], serves as the cornerstone of NMT. It consists of an encoder and a decoder (presented in section 3.4.3.3), both of which can be implemented as recurrent neural networks (RNN), long short-term memory (LSTM), or gated recurrent unit (GRU) models (presented in section 3.4.3.1). The encoder reads the input sequence and condenses the information into a context vector, which is passed on to the decoder to make accurate predictions. The decoder uses the context vector as its initial state, and generates the output sequence, taking into account its own previous outputs as well.

Usually, word embeddings (presented in section 3.4.3.2) serve as a foundational component in the encoder and decoder, converting words into a continuous vector representation. These embeddings encapsulate the semantic meaning of words, thereby facilitating the NMT model's comprehension of the interconnections between the words in the input and output sequences.

Despite the widespread use of the seq2seq architecture in state-of-the-art NMT models, the compression of the source sentence into a single vector by the encoder can result in information loss and difficulties for the decoder in extracting relevant information. To address this problem, the attention mechanism (presented in section 3.4.3.4) allows the decoder to weigh different parts of the source sentence and make informed decisions at each step of the target sentence generation. This results in a more nuanced representation of the source sentence instead of a fixed representation of the entire source sentence. The attention mechanism thus helps to mitigate the issue of incorrect translations produced due to the limitations of fixed representations.

3.4.3.1 Recurrent Neural Network

RNNs are a type of artificial neural network that are specifically designed to handle sequential data. Unlike feed-forward neural networks, where data flows in one direction from input to output, RNNs have a feedback loop, allowing information to be passed from one step of the sequence to the next. This allows RNNs to retain information about past inputs, making them well-suited for tasks such as language modeling, speech recognition, and sequential decision-making. Figure 3.2 explains the working concept of RNNs.

As seen in the Figure 3.2:

- x_t is the input data in the time t , so x_{t-1} is the input data in time $t - 1$.
- y_t is the output in time t and y_{t-1} is the output data in time $t - 1$.

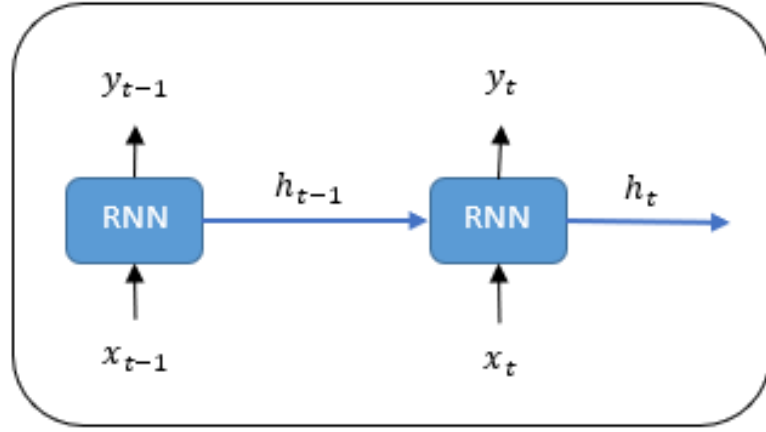


Figure 3.2: Illustrating the Functioning of RNNs for Processing Sequential Data

- h_t is the information produced by the RNN at time t , representing a hidden state of the input at time t .
- h_{t-1} is the hidden state of the input x_{t-1} .

RNNs use the previous hidden state h_{t-1} , to calculate the hidden state h_t using the equation 3.3. This updated hidden state h_t is then used in equation 3.4 to produce the output y_t at time t .

$$h_t = \tanh(w_{hh}h_{t-1} + w_{hx}x_t) \quad (3.3)$$

$$y_t = \text{softmax}(w_{hy}h_t) \quad (3.4)$$

Where w_{hh} , w_{hx} , and w_{hy} are the internal weights at time t .

The training of neural networks involves three key steps:

- Forward propagation, where the network makes predictions by allowing information to flow from the input nodes, through the hidden layers, to the output nodes.
- The calculation of a loss function, which measures the difference between the network's prediction and the actual output, producing an error value.
- Backpropagation, which uses the error value to determine the gradients of the nodes in the network. These gradients are then used to update the network's internal weights, leading to improved performance.

RNNs utilize a backpropagation technique known as backpropagation through time. The adjustment of internal weights is directly proportional to the increase or decrease in gradients. In the process of backpropagation, each node computes its gradient based on

the gradients from the preceding layer. If the adjustments in the previous layer decrease, the adjustments in the current layer will decrease even further, resulting in an exponential decrease of gradients as backpropagation continues down the network. This decrease can result in the vanishing gradient problem [53, 54], which impairs the adjustment of weights and negatively impacts the learning ability of the network.

However, as the gradients propagate through each time step, they become exponentially smaller. This leads to the vanishing gradients problem, causing the RNN to struggle with learning long-range dependencies across time steps, leading to a short-term memory problem. To address this, solutions such as long short-term memory (LSTM) and gated recurrent units (GRU) have been developed.

Long Short-Term Memory

The LSTM [55] network is designed to overcome the limitations of traditional RNNs with the use of a memory cell state. The memory cell state acts as a means of transporting relevant information throughout the sequence chain.

In LSTM, the current hidden state h_t , is computed using the formula 3.5. The cell state c_t , is computed using formula 3.6 and is influenced by three gates, forget gate f_t , input gate i_t , and output gate o_t , which are neural networks that determine which information is allowed to pass through the cell state.

$$h_t = o_t * \tanh(c_t) \quad (3.5)$$

$$c_t = f_t * c_{t-1} + i_t * \tanh(w_{hf}h_{t-1} + w_{xf}x_t) \quad (3.6)$$

The forget gate f_t (computed using the formula 3.7) uses information from both the previous hidden state h_{t-1} and the current input x_t to decide what information should be forgotten or kept, with values ranging from 0 to 1, where 0 means to forget and 1 means to keep.

$$f_t = \sigma(w_{hf}h_{t-1} + w_{xf}x_t) \quad (3.7)$$

The input gate i_t (computed using the formula 3.8) updates the cell state by transforming the values to be between 0 (not important to keep) and 1 (important to keep) using the previous hidden state h_{t-1} and the current input x_t .

$$i_t = \sigma(w_{hi}h_{t-1} + w_{xi}x_t) \quad (3.8)$$

The output gate o_t (computed using the formula 3.9) determines the information that is passed on to the current hidden state h_t and the current output y_t .

$$o_t = \sigma(w_{ho}h_{t-1} + w_{xo}x_t) \quad (3.9)$$

The current output y_t is computed using formula 3.4, and the previous long-term memory c_{t-1} at time $t - 1$ is used as input. The sigmoid function σ , is computed using formula 3.10.

$$f(n) = \frac{1}{1 + e^{-n}} \quad (3.10)$$

Gated Recurrent Unit

The GRU was developed by [56] to enhance each recurrent unit's capability of recognizing dependencies across different time scales, similar to the LSTM unit. Unlike the LSTM, the GRU doesn't have a standalone memory cell and instead uses gating mechanisms to manage the flow of information within the unit [57].

The GRU computes a candidate hidden state h at each time step using the previous hidden state h_{t-1} and the current input x_t , and updates the hidden state based on the reset and update gates (formula 3.11).

$$h = \tanh(W_h * [r_t * h_{t-1}, x_t] + b_h) \quad (3.11)$$

Where W_h and b_h are the weight matrix and bias for the candidate hidden state, respectively. The Update Gate z_t is computed using the formula 3.12 where W_z and b_z are the weight matrix and bias for the update gate, respectively, and σ is the sigmoid activation function. The sigmoid function maps the input to the range $[0,1]$, and the update gate determines what information to keep from the previous hidden state and what new information to add from the current input.

$$z_t = \sigma(w_{hz}h_{t-1} + w_{xz}x_t) \quad (3.12)$$

The Reset Gate r_t is computed using the formula 3.13, where W_r and b_r are the weight matrix and bias for the reset gate, respectively. The reset gate decides how much of the past information to forget.

$$r_t = \sigma(w_{hr}h_{t-1} + w_{xr}x_t) \quad (3.13)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h \quad (3.14)$$

Then, Current Hidden State h_t is computed using the formula 3.14, it is calculated by combining the previous hidden state and the candidate hidden state, weighted by the update gate value. The weights W_r , W_z , W_h and biases b_r , b_z , b_h are learned during the training

process and control the behavior of the GRU cell.

3.4.3.2 Word Embeddings

The fundamental building block in NMT is either an RNN, LSTM, or GRU cell, which are mathematical functions. To input data into these cells, it must be numerical. However, in the context of NMT, the input data is text, not numbers. Thus, the text must first be converted into numerical representations, known as word indices, before it can be processed by these cells.

Let's consider a sentence X with the length l , where each token in X is represented by x_i , with x_i being a member of the vocabulary Σ . X can be written as:

$$X = x_1^l = x_1, \dots, x_l \quad (3.15)$$

When assigning a token index to each x_i , it is important to ensure that each token has a unique index. One-hot vectors are a simple way to represent tokens, where the i -th token x_i in the vocabulary Σ is represented by a vector with a 1 on the i -th dimension to indicate its presence and 0 on the rest of the dimensions to indicate its absence. This is a basic way to describe categorical features in Machine Learning. However, one-hot vectors are not the best representation for tokens in NMT because:

1. The dimensionality of the vectors is equal to the vocabulary size, leading to lengthy vectors if the vocabulary is large, which is not ideal in practice.
2. One-hot vectors do not capture the meaning of the words.

Word embeddings address the limitations of one hot encoding by representing words as dense vectors of continuous numbers, rather than discrete and sparse representations. This representation is learned through unsupervised training on a large corpus of text, where the model identifies patterns and relationships between words. The resulting embeddings capture the semantic meaning and similarity between words and represent them in a continuous numerical space, where the distance between words reflects their relatedness.

3.4.3.3 Sequence-to-Sequence

The Encoder-Decoder architecture forms the basis of the Recurrent Continuous Translation Models (RCTM) I and RCTM II that were developed by Kalchbrenner and Blunsom [45]. This family of networks plays a crucial role in the translation of source sentences into target sentences. The process begins with the source sentence being passed to an encoder, which encodes all the information in the source sequence into a single real-valued vector known as the context vector. The context vector is then passed to the decoder, which generates the target language output sequence.

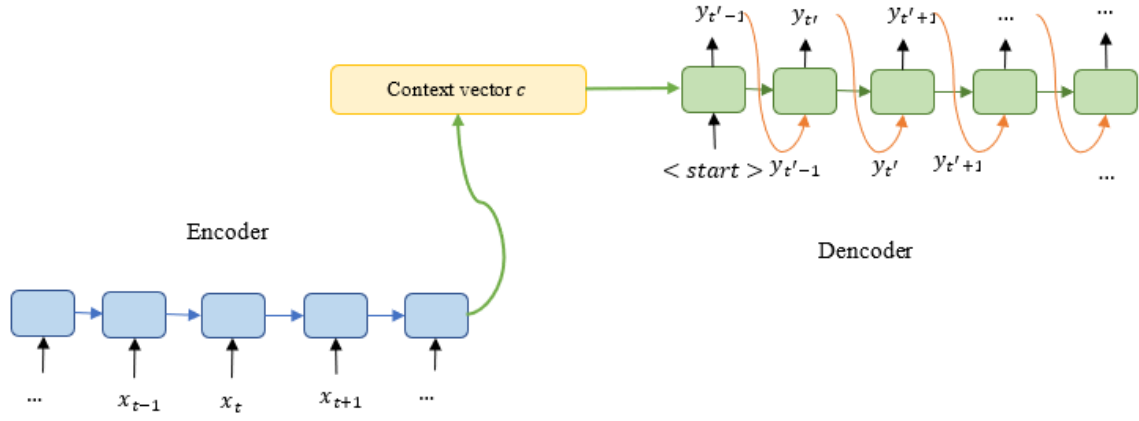


Figure 3.3: Seq2seq NMT Model Architecture.

The encoder starts by converting the source sentence into word representations that collect semantic knowledge about the words. These representations are referred to as encoder representations. Both the encoder and decoder are composed of a set of RNN, LSTM, or GRU cells, and the general concept of the encoder can be seen in a Figure 3.3.

Typically, the encoder transforms an input sequence x of variable length into a fixed-shape context variable c . At time t , the RNN cell transforms the input feature vector x_t and the previous hidden state h_{t-1} into the current hidden state h_t . Therefore, each RNN cell uses the hidden state output of the preceding RNN cell to generate its own hidden state. Through a custom function f , the encoder transforms the hidden states h into the context vector c at all time steps, as shown in the Figure 3.3, instead of only converting the last hidden state to a context vector.

If the encoder uses GRU cells, the context vector is the hidden state h . However, if the encoder uses LSTM cells, the context vector is both the hidden state h and cell state c . The encoder representations are used by the decoder to generate the target sentence, as depicted in the Figure 3.4. The encoder uses the context vector c generated by the decoder to encode the entire input sequence $\{x_1, \dots, x_t\}$ and produce the output sequence $\{y_1, \dots, y_{t'}\}$ from the training dataset.

For each time t' , where t' is different from t , the probability (as defined in formula 3.16) of decoding the output $y_{t'}$ is dependent on the previous output subsequence $\{y_1, \dots, y_{t'-1}\}$ and the context variable c . The decoder uses another RNN to generate the current hidden state $s_{t'}$ using the previous hidden state $s_{t'-1}$, previous output $y_{t'-1}$, and the context vector c . The function g (as defined in formula 3.17) expresses the transformation of the decoder's hidden layer.

$$P(y_{t'} | y_1, \dots, y_{t'-1}, c) \quad (3.16)$$

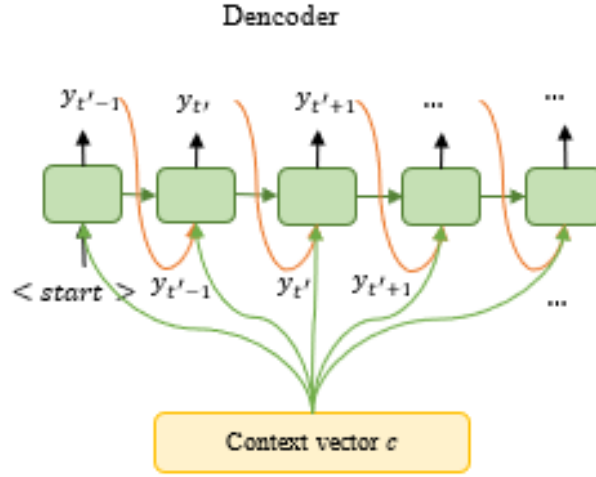


Figure 3.4: Decoder Architecture in Seq2Seq NMT Model: Utilizing Context Vector for Output Prediction.

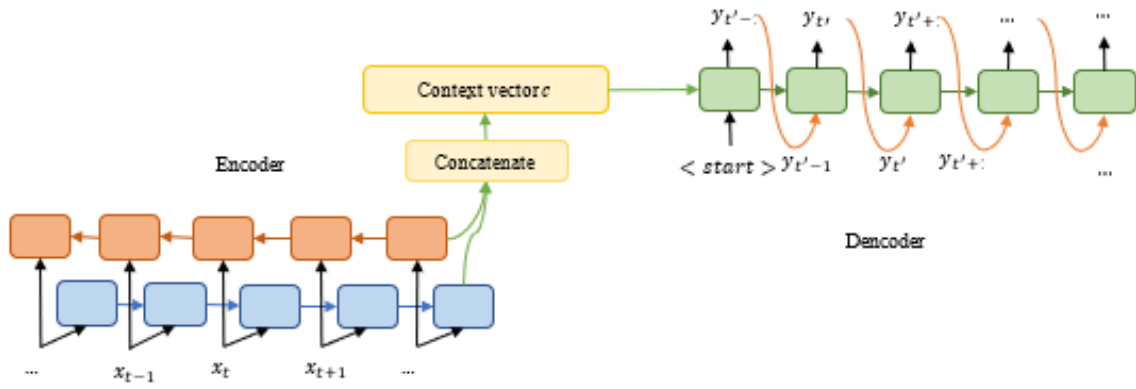


Figure 3.5: Bidirectional Seq2Seq NMT Model Architecture.

$$s_{t'} = g(y_{t'-1}, c, s_{t'-1}) \quad (3.17)$$

The encoder depicted in Figure 3.3 is referred to as a unidirectional encoder, as it only processes information from the past. This means that it only takes inputs from the past and retains this information. On the other hand, the encoder shown in Figure 3.5 is called a bidirectional encoder. It operates by processing inputs in two directions - first from past to future (t to $t + 1$), referred to as the forward direction, similar to the unidirectional encoder. Second, it processes inputs from the future to the past ($t + 1$ to t), known as the backward direction, allowing it to retain information from the future. By combining the two hidden states produced by the forward and backward passes, the bidirectional network is able to preserve information from both the past and the future at any given point in time t .

It is widely accepted in the deep learning community that deeper models tend to pro-

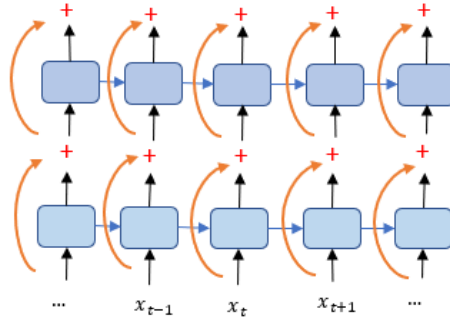


Figure 3.6: Stacked LSTM layers in Seq2seq NMT Model Architecture with Residual connections.

duce improved accuracy results as they perform a more comprehensive analysis of the training data and generate better feature representations. However, it should be noted that adding excessive layers of LSTM may not necessarily result in further accuracy improvements and can cause difficulties in training due to the problems of exploding and vanishing gradients. To tackle these issues, residual connections [51] have been introduced as a solution. The residual connection involves adding the input to the output of an LSTM layer and passing the sum as the input to the next LSTM layer, as illustrated in Figure 3.6. This helps enhance accuracy and mitigate the exploding and vanishing gradient problems.

3.4.3.4 Attention Mechanism

In the encoder-decoder model, fixed context vector representation is suboptimal due to:

- The encoder cannot encode the entire information about the long sentences.
- The decoder can utilize only one representation of the source during decoding. However, distinct segments of the source carry varying degrees of importance at each generation step. Furthermore, starting from the current state, the decoder faces challenges in extracting all the pertinent information solely from a fixed representation.

The performance of NMT models has improved with the addition of the attention mechanism. The attention mechanism is a component of a neural network that helps to improve translation quality.

At each generation step, the decoder is able to determine which parts of the source are most relevant, while the encoder no longer needs to condense the entire source into one vector. Instead, the encoder provides individual representations for every source token. The attention mechanism then enables the network to determine the importance of each input part at each step. To do this, the attention mechanism uses the decoder's state h_t and all encoder's state s_k as input and calculates attention scores $score(h_t, s_k)$.

$score(h_t, s_k)$ is calculated in three different ways, dot-product (formula 3.18), bilinear function (formula 3.19), and multi-layer perception (formula 3.20) methods, where w, w_2^T , and w_1 are the internal weights.

$$score(h_t, s_k) = h_t s_k \quad (3.18)$$

$$score(h_t, s_k) = h_t^T w s_k \quad (3.19)$$

$$score(h_t, s_k) = w_2^T \cdot \tanh(w_1 [h_t, s_k]) \quad (3.20)$$

The attention scores $score(h_t, s_k)$ are used to compute the attention weights $a_k^{(t)}$. The *SoftMax* function is applied to the attention scores to calculate the attention weights, as shown in formula 3.21. The attention output c_t is then computed by taking the weighted sum of the encoder states s_k with the attention weights, as shown in formula 3.22.

$$a_k^{(t)} = \frac{\exp(score(h_t, s_k))}{\sum_{i=1}^m \exp(score(h_t, s_k))}, k = 1..m \quad (3.21)$$

$$c_t = \sum a_k^{(t)} s_k \quad (3.22)$$

In this context, two models were developed, Bahdanau attention model [48] and Loung attention model [58].

Bahdanau Attention Model

The Bahdanau Model adheres to the encoder-decoder framework and utilizes bidirectional RNNs in the encoder to concatenate the states of the two RNNs for each token. This process results in a better encoding of each source word, allowing for the capture of complex dependencies in the input. To compute attention scores, the model applies a multi-layer perceptron (MLP) (Equation 3.20) between the encoder and decoder, generating an output vector that is passed to the decoder at each step. The attention mechanism is used between decoder steps and involves the use of the previous decoder state to compute attention and generate an output vector, which is then passed to the decoder at the current step along with the previous decoder state. This iterative process continues until the model generates an end-of-sentence token, completing the translation.

Luong Attention Model

The Luong Model also followed the encoder-decoder framework. The encoder in this model is a simple unidirectional RNN that reads the input sequence in a single direction.

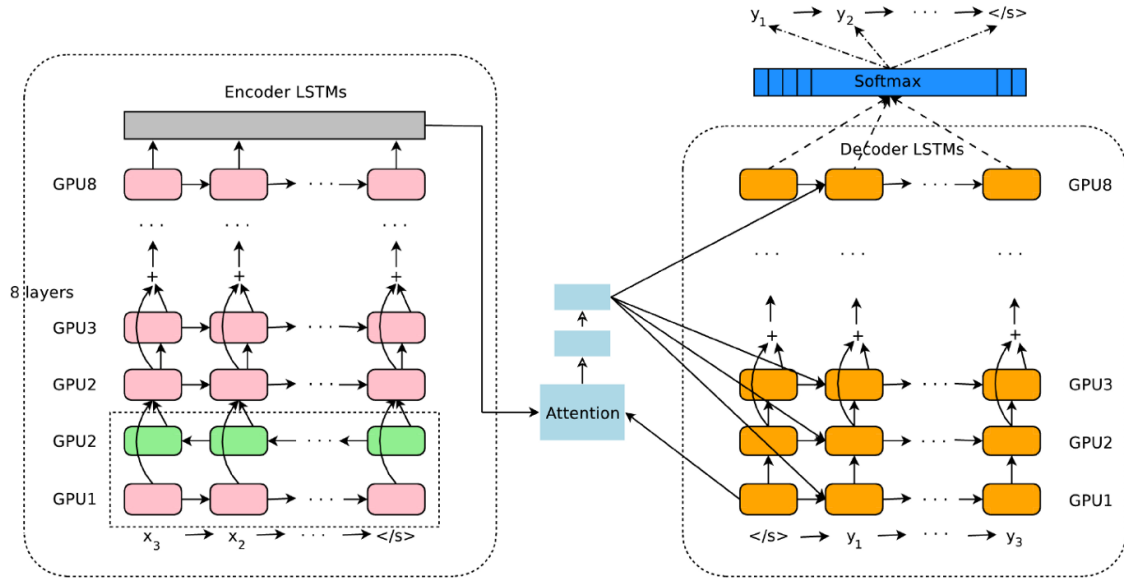


Figure 3.7: The GNMT Architecture [51].

The attention scores between the encoder and decoder are computed using a bilinear function (Equation 3.19). The attention mechanism is applied between the decoder RNN state and the prediction for that specific step. After the RNN decoder step, the attention mechanism is used to compute the output state, which is then combined with the decoder RNN state to obtain an updated representation. This updated representation is then used to generate the prediction.

3.4.3.5 Fundamental Architecture in NMT

Unfortunately, NMT systems are infamous for being computationally expensive in both training and translation inference, as well as for lacking robustness, especially when input sentences contain rare words. These limitations have hampered the use of NMT in actual deployments and services where accuracy and speed are important. In this section, we present the fundamental NMT models, which attempt to address many of these issues.

Based on seq2seq architecture and attentional mechanism, there are four fundamental architectures proposed, Google's Neural Machine Translation system (GNMT) as an example of a recurrent network, the convolutional sequence to sequence model (ConvS2S), the self-attention-based Transformer model, and an interesting fusion is the RNMT+ model.

Google's Neural Machine Translation System

GNMT architecture [51] follows the attentional seq2seq framework, when its encoder and its decoder are connected through an attention module which allows the decoder to focus on different regions of the source sentence during the course of decoding.

The GNMT model architecture is depicted in Figure 3.7, where the encoder is positioned on the left, the decoder on the right, and the attention module in the middle.

To expedite training, the model is divided into eight GPUs, with the SoftMax layer at the end responsible for predicting the final output. Each GPU contains an encoder layer and its corresponding decoder layer, enabling parallel processing during training.

The GNMT encoder comprises a bi-directional LSTM layer followed by seven uni-directional LSTM layers. The decoder is formed by eight uni-directional LSTM layers, with a SoftMax layer positioned at the top. In the middle of the architecture, a single attention mechanism is located. Both the encoder and decoder utilize residual skip connections between consecutive levels.

In the encoder, the bidirectional LSTM layers operate in such a way that the lower LSTM nodes gather information from left to right, while the upper LSTM nodes gather information from right to left. The outputs of these bidirectional layers are then concatenated and passed as input to the next layer. To address the challenges associated with training deep models, the subsequent unidirectional LSTM layers in the encoder are interconnected using residual connections. These residual connections involve adding the output of $layer_t$ to the output of $layer_t - 1$, resulting in a final addition that serves as input to $layer_{t+1}$. This approach helps mitigate the difficulties and slow training that arise when stacking a large number of LSTM layers due to issues such as exploding and vanishing gradients.

Moreover, all layers contained in the decoder network are uni-directional LSTM layers, the first and the second layers are simply connected with each other, while, such as in the encoder, the six last layers in the decoder are connected with each other with the residual connection.

In addition, the attention module used is proposed in [48], the encoder output and the output of the first layer of the decoder are given as input to the attention mechanism, and the output of the attention mechanism will be provided as input in all next decoder layers.

The Convolutional Sequence-to-Sequence

ConvS2S model [59] is based on the convolutional neural network (CNN), where convolutional neural networks are used for both the encoder and the decoder instead of RNNs to compute the intermediate encoder and decoder states. Figure 3.8 shows the ConvS2S architecture, where the encoder and decoder composed a multiple block, we refer the th block by l , where each block is devised into two sub-blocks, the first was one-dimensional convolutional layer, and the second was gated non-linear unit. The last decoder block is followed by multi step attention, which is dot products attention between decoder context representations and encoder representations, then, the attention output is added to the decoder states which then predict the target words.

The one-dimensional convolutional layer is used to computes intermediate states based

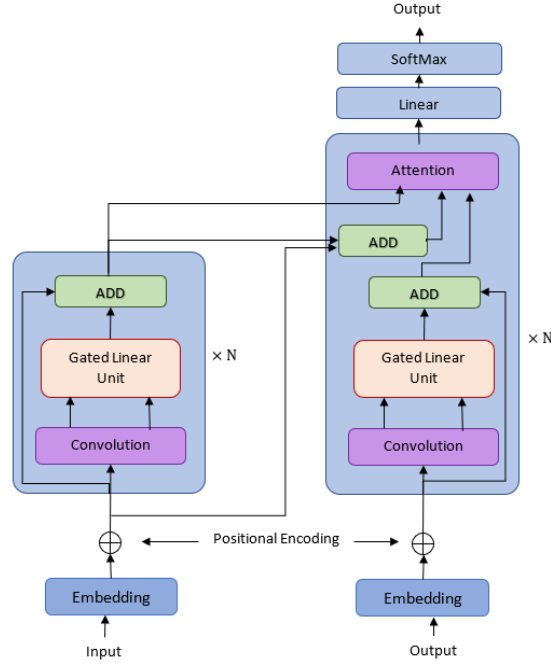


Figure 3.8: The Convs2s Architecture [59].

on a fixed number of input elements. For each network, encoder or decoder, the output of encoder l block is $z^l = (z_1^l, \dots, z_m^l)$, the output of decoder l block is $h^l = (h_1^l, \dots, h_n^l)$, and k is the kernel width of l block. Each output state h_1^l contains information over k input elements, and the number of input elements represented in a state increase by stacking several blocks on top of each other.

Each convolution kernel has two parameters: weight $W \in \mathbb{R}^{2d \times kd}$ and bias $b_w \in \mathbb{R}^{2d}$. Its input, $X \in \mathbb{R}^{k \times d}$, is a concatenation of k input elements embedded in d dimensions, and it maps those elements to one output element, $Y \in \mathbb{R}^{2d}$, which has twice as many dimensions as the input elements. The k output elements from the current layers are used by the next layers, and so forth.

Non-linearity is implemented through Gated Linear Units, which employ a simple gating mechanism over the output of the convolution $Y[AB] \in \mathbb{R}^{2d}$:

$$v([AB]) = A \otimes \sigma(B) \quad (3.23)$$

Where:

- $A, B \in \mathbb{R}^{2d}$ are the inputs to the non-linearity.
- $\otimes$ is the point-wise multiplication.
- The output $Y[AB] \in \mathbb{R}^{2d}$ is half the size of Y .

The gates $\sigma(B)$ control which inputs A of the current context are relevant. Non-linearities gives the networks the ability to utilize the whole input elements or, if necessary,

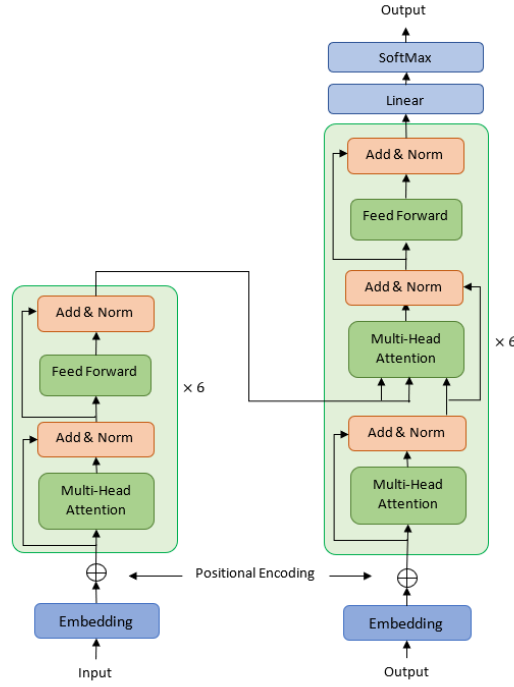


Figure 3.9: The Transformer Architecture [60].

to focus on fewer elements.

Furthermore, residual connections are added from the input of each convolution to the output of the block to enable deep convolutional networks. In addition, Position embeddings are used to inform the model about which segment of the sequence in the input or output it is currently dealing with. And at the top of the network, a SoftMax layer is put in to predict the final output.

Transformer

Such as GNMT and ConvS2S models, the transformer [60] also follows the encoder decoder framework. However, it does not rely on recurrence and convolutions in order to generate an output. As a result, the Transformer model cannot capture information about the relative positions of the words in the sequence intrinsically, and as a solution to this problem, positional encodings are added to the input embeddings in order to inject this information. Each encoder and decoder consist of a stack of $N = 6$ identical layers as shown in Figure 3.9.

The six encoder layers, each of them composed of two sublayers:

- The first sublayer implements a multi-head self-attention mechanism, which includes heads that receive a linearly projected version of the queries, keys, and values and generate outputs in parallel, which are then utilized to generate a final result.
- The second sublayer is a fully connected feed-forward network made up of two linear

transformations separated by Rectified Linear Unit (ReLU) activation.

On the opposite side, the six decoder layers, each of them composed of three sublayers:

- The first sublayer gets the decoder stack's previous output, augments it with positional information, then applies multi-head self-attention to it. While the encoder is meant to attend to all words in the input sequence regardless of their position, the decoder is adjusted to only attend to the words that come before them. As a result, the prediction for a word at the position can only be based on the known outputs for the words before it in the sequence.
- The second sublayer additionally includes a multi-head self-attention mechanism that takes queries from the previous decoder sublayer as well as keys and values from the encoder output, allowing the decoder to attend to all of the words in the input sequence.
- And, the third sublayer implements a fully connected feed-forward network.

Furthermore, each of these two sublayers is surrounded by a residual connection and succeeded by a normalization layer [61], which normalizes the sum computed between the sublayer input of the sublayer x , and the output generated by the sublayer itself (equation 3.24).

$$layerNorm(x + Sublayer(x)) \quad (3.24)$$

The transformer model addresses to the following problems in the previous two models:

- Unlike GNMT, to maximize parallelization, the Transformer model removes any sequential dependencies in both the encoder and decoder networks.
- The Transformer model makes extensive use of self-attention networks [62] to ensure that each location in the current layer gets access to data from all other positions in the preceding layer in order to solve the restricted context problem that exists in ConvS2S.

Multi-head attention also called self-attention (as shown in Figure 3.10) is one of the key components of the model, it allows the model to jointly attend to information from different representation subspaces at different positions.

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^o \quad (3.25)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3.26)$$

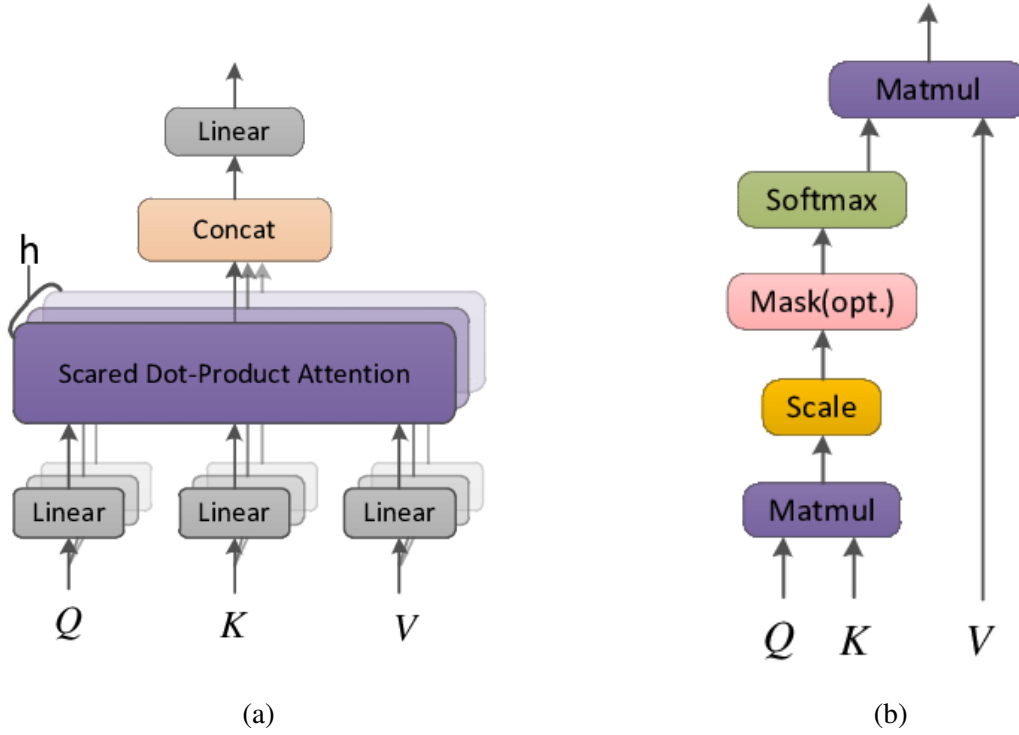


Figure 3.10: Attention Mechanisms in the Transformer. (a) Multi-Head Attention consists of several attention layers running in parallel, and (b) Scaled Dot-Product Attention [60].

Where a head is Attention function may be defined as a function that maps a query and a set of key-value pairs to an output. Where the query, key and value representations roles are asking information, requesting for information, indicates the presence of information, and providing the information, respectively, where the query, key, value and output are all vectors. The output is a weighted sum of the values, with the weight assigned to each value calculated using a Scaled Dot-Product function of the query with the appropriate key, which is given by the equation. Q, K of dimension d_k , and V of dimension d_v , and parameter matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_k}$ and $W^O \in \mathbb{R}^{d_{model} \times d_k}$, where, $d_k = d_v = d_{model}$ and $h = 64$. The transformer used in each multi-head attention eight head.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (3.27)$$

In the transformer, as we mentioned above, each layer in the encoder or the decoder contains two sub-layers, one is multi-head attention, and the other is a fully connected feed-forward network, which is applied to each position separately and identically. A fully connected feed-forward network is composed of two linear transformations separated by a ReLU activation (equation 3.28), while the linear transformations are the same across all positions, the parameters weight (W_1, W_2) and bias (b_1, b_2) are vary from layer to layer, with the dimensionality of input and output is $d_{model} = 512$.

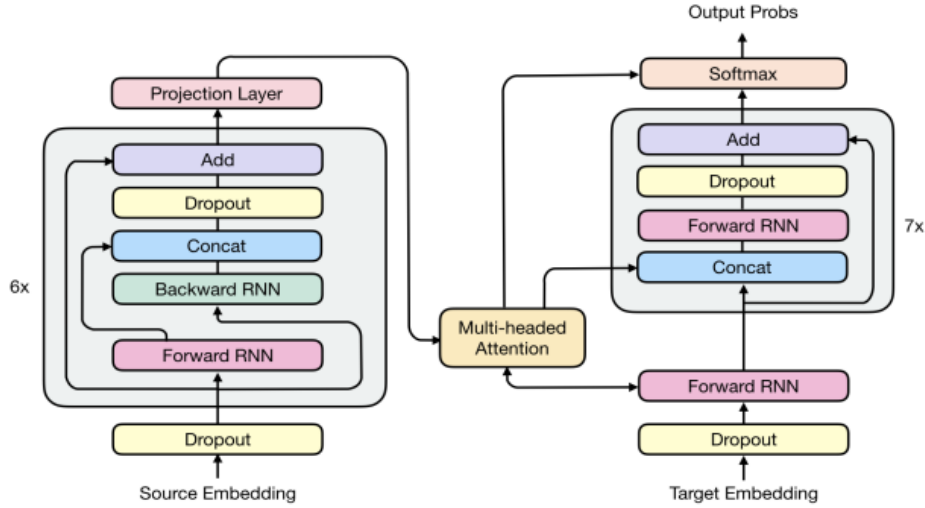


Figure 3.11: Model architecture of RNMT+ [63].

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2 \quad (3.28)$$

RNMT+

The newly proposed architecture in the field of NMT is RNMT+ [63]. It builds upon previous models such as the GNMT and the Transformer, and introduces several innovative features to improve translation quality. The architecture features a multi-layered encoder and decoder with non-identical layers between them, as depicted in Figure 3.11.

The encoder in RNMT+ has six stack layers, each of which consists of a bidirectional LSTM layer followed by a dropout layer. Bidirectional LSTMs process the input sequence in both forward and backward directions, which allows them to capture contextual information from both past and future tokens. Dropout is a regularization technique that randomly drops out some of the units in the layer during training, preventing overfitting and improving generalization.

At the top of the encoder layers, a projection layer is added to map the hidden states to a higher-dimensional space. This is done to increase the representational power of the model and allow for better information flow between the encoder and decoder.

The decoder in RNMT+ has seven stack layers, each layer consists of a unidirectional LSTM followed by a multi-head additive attention mechanism and a normalization layer. The normalization layer [61] helps to stabilize the gradients during training and improve the overall performance of the model.

In addition, each decoder layer is followed by a residual connection, which allows the gradient to flow more easily through the network and prevents the vanishing gradient problem. The residual connection also helps to preserve important information from earlier

layers and avoid the problem of information loss.

To compute the attention context, the multi-head additive attention mechanism is applied between the output of the final encoder layer and the input to the first decoder layer. This produces a set of context vectors, which are then concatenated with the input to each decoder layer.

Finally, the output of the last decoder layer is passed through a SoftMax layer to obtain the probability distribution over the target language vocabulary. During training, the model is optimized to minimize the cross-entropy loss between the predicted and actual target sequence.

3.5 Metric Evaluation of Machine Translation

The success of a MT system is highly dependent on the quality of its output. Traditionally, there are several methods to evaluate the system's output quality, such as human evaluations, reading tests, and other related assessments. Human evaluations, which use bilingual or monolingual judges, are the most commonly used approach to assess MT output. Bilingual judges evaluate the output's accuracy and completeness in conveying the source information, while monolingual judges assess the fluency and naturalness of the output in the TL [64, 65]. Despite their widespread use, human evaluations can be subjective, time-consuming, expensive, and lack repeatability [66].

To address these challenges, automated metrics have been developed to evaluate the quality of MT output. BLEU and METEOR scores are two of the most commonly used metrics in MT evaluation. These metrics work by comparing the system output to a set of reference translations of the source data, and determining the extent to which the system output matches the references. This provides a more objective, efficient, and repeatable way of evaluating the quality of MT output.

3.5.1 BLEU Score

Bilingual Evaluation Understudy (BLEU) [67] is a metric that evaluates the accuracy of MT systems by focusing on precision. Rather than evaluating the degree to which the references are replicated in the system output, BLEU score determines the correctness of the output. Since its introduction, BLEU score has become the standard evaluation metric in the MT community [68].

To calculate the BLEU score of a system's output, the overlap between the output and a set of reference translations is measured in terms of the number of matching n-grams, or sequences of words. This information is then used to calculate the precision of the system's output, which forms the basis of the BLEU score.

Formula 3.29 illustrates the calculation of the BLEU precision scores. The precision score, represented by p_n , is determined by adding up the clipped n-gram counts for all candidate sentences in the corpus and dividing by the total number of candidate n-grams.

$$p_n = \frac{\sum_{C \in \{\text{candidates}\}} \sum_{n\text{-gram} \in C} \text{Count}_{\text{clip}}(n\text{-gram})}{\sum_{C' \in \{\text{candidates}\}} \sum_{n\text{-gram}' \in C'} \text{Count}_{\text{clip}}(n\text{-gram}')} \quad (3.29)$$

The next step in the BLEU evaluation process involves calculating the BLEU brevity penalty, represented by BP . This is done using Formula 3.30, where c is the length of the candidate translation and r represents the effective length of the reference corpus.

The BP and p_n values are then combined to compute the final BLEU score, as outlined in Formula 3.31. The parameter N is typically set to 4, and the weight w_n is usually assigned a value of $1/N$. This results in the overall BLEU score, which provides a measure of the accuracy of the system's output.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{(1 - \frac{r}{c})} & \text{if } c < r \end{cases} \quad (3.30)$$

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right) \quad (3.31)$$

3.5.2 METEOR Score

Metric for Evaluation of Translation with Explicit Ordering (METEOR) [69] is a recall-oriented evaluation metric that addresses the limitations of BLEU score. Unlike BLEU score, which only considers precision, METEOR score calculates both precision and recall and integrates them into one score.

The METEOR score is calculated using the following formula:

$$\text{score} = (1 - \text{Pen}) \cdot F_{\text{means}} \quad (3.32)$$

The fragmentation penalty (Pen) is applied to bias the METEOR score against system outputs that have many short, consecutive matching sequences known as chunks. This penalty is calculated as the number of chunks divided by the number of unigrams matches, using formula 3.33 with default values of $\beta = 3$ and $\gamma = 5$. And, the harmonic mean, represented by the symbol F_{means} , is calculated using formula 3.34.

In addition, the precision (P) in METEOR is defined as the ratio of the number of matching words to the total number of words in the system output. And, Recall (R) is defined as the ratio of the number of matching words to the total number of words in the reference.

$$\text{Pen} = \gamma \text{Frag}^\beta \quad (3.33)$$

$$F_{means} = \frac{P.R}{\alpha P + (1 - \alpha).R} \quad (3.34)$$

The word matching process in METEOR score involves several stages, each focused on aligning strings in the system output and reference translations. During each stage, only words that were not previously matched are considered for matching. The comparison is limited to individual words (unigrams). The stages are:

1. Exact Matching: Identical strings in the reference and hypothesis are aligned.
2. Stem Matching: Stemming is applied so that words with the same morphological root are aligned.
3. Synonymy Matching: Words that are synonyms according to WordNet [70] are aligned.

3.6 Conclusion

In this chapter of our thesis, we delve deeper into the field of MT by examining how a machine can understand and translate a sentence just like a human. We begin by exploring the various techniques used in MT, starting with the oldest method, RBMT. We then describe the three sub-methods involved in this technique, Direct, Transfer, and interlingua MT.

Following this, we move on to CBMT, which includes three sub-methods: SMT, EBMT, and NMT. In particular, we delve into NMT, discussing key components such as word embeddings, seq2seq models, attention mechanisms, the fundamental elements of neural machine translation, including RNNs, LSTMs, and GRUs, and the fundamental architecture proposed in NMT.

Finally, we touch on the topic of evaluating MT, focusing on the three most commonly used metrics, BLEU and METEOR scores, which provide a way to automatically measure the accuracy and effectiveness of a machine-generated translation.

Overall, this chapter provides a comprehensive overview of the field of MT, offering insights into the different techniques used and the challenges and opportunities they present.

4

Arabic Dialect Neural Machine Translation Overview and Contribution

4.1 Introduction

NMT has become an important research area in NLP in recent years. It has revolutionized the field of MT by achieving state-of-the-art performance on various language pairs. In this chapter, we provide an overview of NMT for the Arabic dialect. Furthermore, we aim to highlight our contributions to NMT within the context of Arabic dialect translation.

In section 4.2, we discuss the Arabic Dialect NMT dataset. We explore the different datasets available for Arabic dialect translation.

Section 4.3 of this chapter focuses on the Arabic dialect NMT and provides an overview of the work in this context. We explore the challenges of Arabic dialect translation and discuss the existing approaches and models used to overcome these challenges.

Section 4.4 is the most important section of this chapter and presents our contribution to the field of Arabic dialect translation using seq2seq, attentional seq2seq, and Transformer NMT models. We present our methods, experimentations, results, and a discussion. Specifically, we discuss our work on Algerian and Arabic dialects translation and the performance achieved by our models. We also compare our results with existing approaches. Finally, section 4.5 presents a conclusion to this chapter.

In summary, this chapter provides an overview of NMT for the Arabic dialect, discussing relevant corpora and the works applied in this area. It also focuses on the specific

contribution made towards NMT in the context of Arabic dialect translation.

4.2 Arabic Dialects Parallel Corpora

The task of MT Arabic dialects, including Algerian dialects, presents a significant challenge due to the lack of comprehensive parallel corpora. Although MT for MSA has seen significant progress in recent years, translating dialects is still a relatively underexplored area of research. As such, the existing parallel corpora for Arabic dialects to MSA translation are relatively limited in number and scope. In this section, we examine the available parallel corpora of Arabic dialects to MSA for MT .

The Multilingual Arabic Dialect Applications and Resources (MADAR) parallel corpus [71] stands as a profoundly comprehensive parallel corpus within this domain. Covering a wide array of dialects, MADAR encompasses the greatest number of these dialects, spanning numerous Arab countries. For each of these dialects, MADAR comprises 2000 sentences, each accompanied by translations into MSA, English, and French. The dialects included are as follows:

- Tunisia: Tunis and Sfax
- Morocco: Rabat and Fes
- Algeria: Algiers
- Libya: Tripoli and Benghazi
- Egypt: Cairo, Alexandria, and Aswan
- Sudan: Khartoum
- Iraq: Mosul, Baghdad, and Basra
- Yemen: Sana'a
- Qatar: Doha
- Syria: Damascus and Aleppo
- Jordan: Amman and Salt
- Saudi Arabia: Riyadh and Jeddah
- Palestine: Jerusalem
- Lebanon: Beirut
- Oman: Muscat

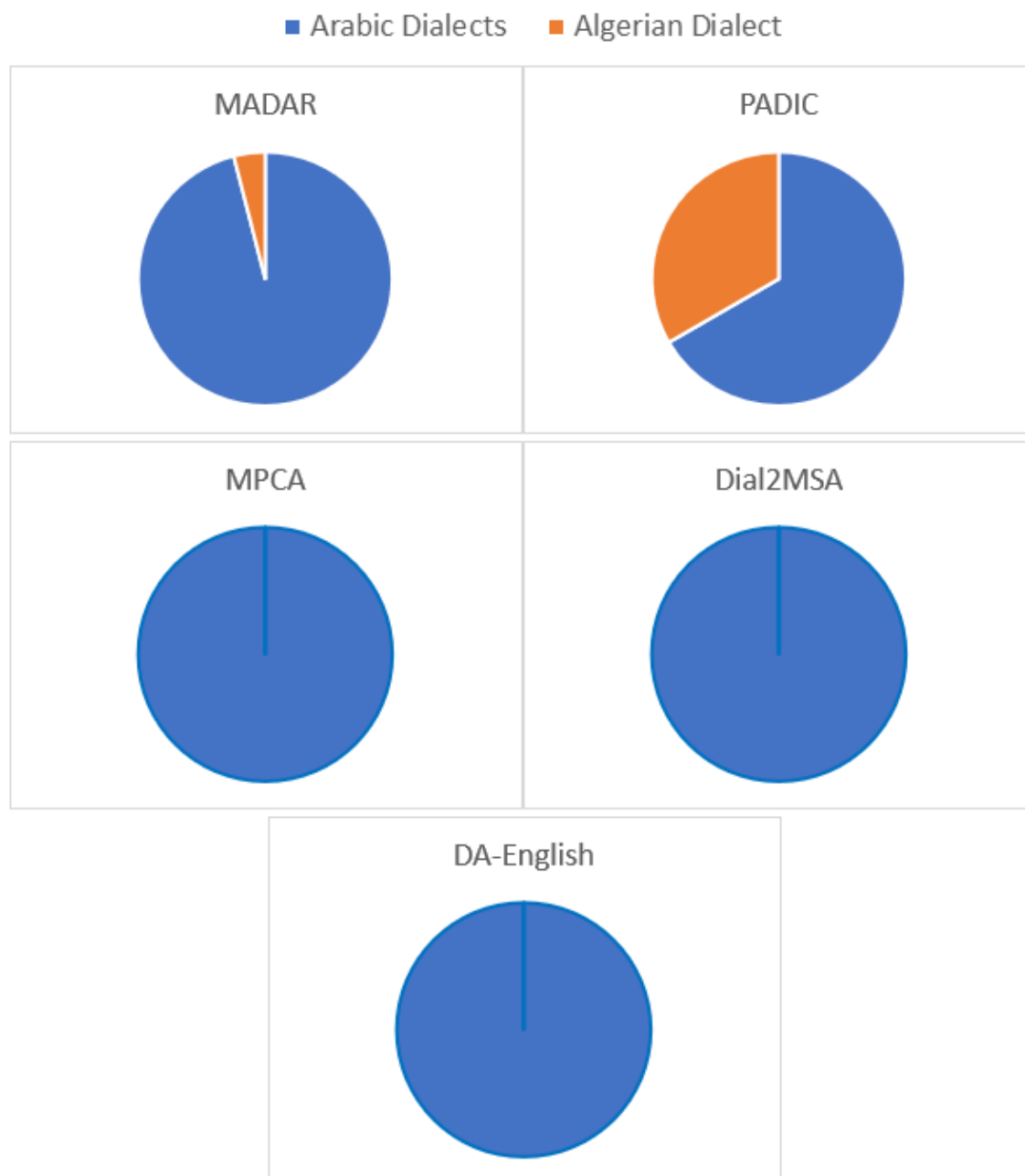


Figure 4.1: Number of Algerian Dialect Sentences in Existing Arabic Dialects parallel corpora. The figure highlights the scarcity of comprehensive parallel corpora containing Algerian dialect sentences among existing Arabic dialects corpora.

Another parallel corpus, the Multidialectal Parallel Corpus of Arabic (MPCA) [72], contains five Arabic dialects, namely Egyptian, Tunisian, Jordanian, Palestinian, and Syrian Arabic dialects. The parallel corpus includes 2000 sentences for each dialect translated into both MSA and English.

The Arabic Parallel Dialects Corpus (PADIC) [3] is another dataset containing five dialects from the Maghreb region (three from Algeria, one from Tunisia, and one from Morocco) and two dialects from the Middle East (Syria and Palestine). Each dialect in PADIC includes 6400 sentences with their MSA translations.

Dial2MSA [73] is a collected corpus from Twitter, containing 16000 Egyptian and 8000 Moroccan dialect sentences with verified MSA translations. The corpus also includes 18000 Levantine and Gulf dialect sentences with MSA translations.

The DA-English Parallel Corpus [5] comprises 38000 Egyptian Arabic sentences translated into English, as well as 138000 Levantine Arabic sentences paired with their English translations. These sentences were gathered from various online user groups and dialectal Arabic weblogs.

This thesis is focused on utilizing NMT to accurately translate Algerian dialects into MSA. However, the primary hurdle in achieving this goal is the lack of comprehensive datasets containing Algerian dialect sentences.

Figure 4.1 shows that out of the existing parallel corpora on Arabic dialects, only MADAR and PADIC include Algerian dialect sentences. However, these parallel corpora have a relatively small number of Algerian dialect sentences, making it challenging to train NMT models effectively. Consequently, the scarcity of data limits the accuracy and effectiveness of MT systems developed for Algerian dialects.

Translating Algerian dialects is a significant challenge due to several factors, such as the diverse range of Algerian dialects with local nuances and variations that must be accounted for. Furthermore, the limited availability of high-quality parallel corpora containing Algerian dialect sentences further complicates this task.

4.3 Arabic Dialects NMT an Overview

The issue of NMT for the Arabic language has been addressed in several studies, including [74, 75] and [76]. [75] outcomes of Arabic NMT were presented for both translation directions. This work involved a comparison between phrase-based systems and an attentional bidirectional seq2seq model. A variety of pre- and post-processing techniques were employed, encompassing morphology-aware tokenization and orthographic normalization. [74] focused on the Amharic-Arabic language pair, comparing two seq2seq models, one built with LSTM cells and the other utilizing CRU cells. Lastly, [76] introduced a character-based NMT model tailored for languages with limited resources. This model amalgamated the strengths of RNN and CNN architectures.

Table 4.1: Related works of Arabic dialect NMT models.

Work	Year	Model	Dataset	Train data size	Language or Dialect	Bleu score
[74]	2019	Attentional LSTM	Seq2seq the authors created a small parallel text corpus by adjusting existing monolingual Arabic	13501 sentences	Amharic-Arabic	12.0
		Attentional GRU	Seq2seq and Amharic text corpora available on Tanzile			10.5
[75]	2016	Attentional bidirectional seq2seq	Combines from LDC2004T18, LDC2004T17 and LDC2007T08.	1.2 M sentences	Arabic-English	49.7
[76]	2018	model combines RNN and CNN	IWSLT2016	1.0246M sentences	Arabic-English	18.67
[77]	2017	Seq2seq	PADIC	12800	Algerian Dialect - MSA	4.3
[78]	2022	Transformer	MADAR and DA-English Parallel Corpus	250.7 K	Egyptian - English	25.72

However, NMT for Arabic dialects is challenging due to the scarcity of high-quality language pair data. Some studies have addressed this challenge. One such study [77] employed seq2seq to directly translate the Algerian dialect to MSA. Another study [78] utilized a transformer model to translate the Egyptian dialect into English. Table 4.1 presents a comparative study between these works.

Developing or enhancing NMT models is a daunting task, particularly due to the paucity of data. Low-resource languages, in particular, have a limited amount of data, making it a major obstacle for researchers and practitioners. Consequently, there are only a few studies available on certain language pairs, such as those involving Arabic dialects. This situation makes it difficult to train more effective NMT models, such as the transformer.

For instance, a study [78] that employed the transformer model for translating Egyptian dialect into English demonstrated that to achieve a BLEU score of 25, the model needed over 250000 sentences for training. Conversely, study [75] that used an attentional seq2seq model for a similar translation task attained a BLEU score of 12.10 with only 13500 sentences, as presented in Table 4.1. This stark contrast in the data required to obtain comparable results underscores the vital role of data availability in creating effective NMT models.

To summarize, having access to high-quality data is crucial for the development and improvement of NMT models. The scarcity of data presents significant challenges to researchers and practitioners in training efficient models. Therefore, there is a pressing need to continue collecting and maintaining quality linguistic data, particularly for low-resource languages. In cases where there is a lack of data, researchers must seek alternative strategies to improve the system's efficiency rather than focusing solely on improving the model to advance the field of MT.

4.4 NMT Models for Arabic Dialects a Contribution

Despite the challenges posed by Arabic dialects, the development of NMT systems for these dialects is an important area of research that has the potential to make a significant impact on communication and understanding in the Arabic-speaking world. By training NMT models on dialect-specific data, it is possible to improve the accuracy and appropriateness of MT for Arabic dialects. In the preceding sections, it was apparent that a requirement exists for dialect translation. However, managing this requirement is made complex by the insufficient availability of data.

To address this challenge, our research investigated the use of seq2seq (Section 4.4.1), attentional seq2seq (Section 4.4.2), and transformer (Section 4.4.3) NMT models for translating from the Algerian dialect to MSA, as seen in Figure 4.2. Expanding further, the same models were also employed for the translation of a majority of other Arabic dialects.

The Arabic language has a diverse range of dialects, each with its own linguistic fea-

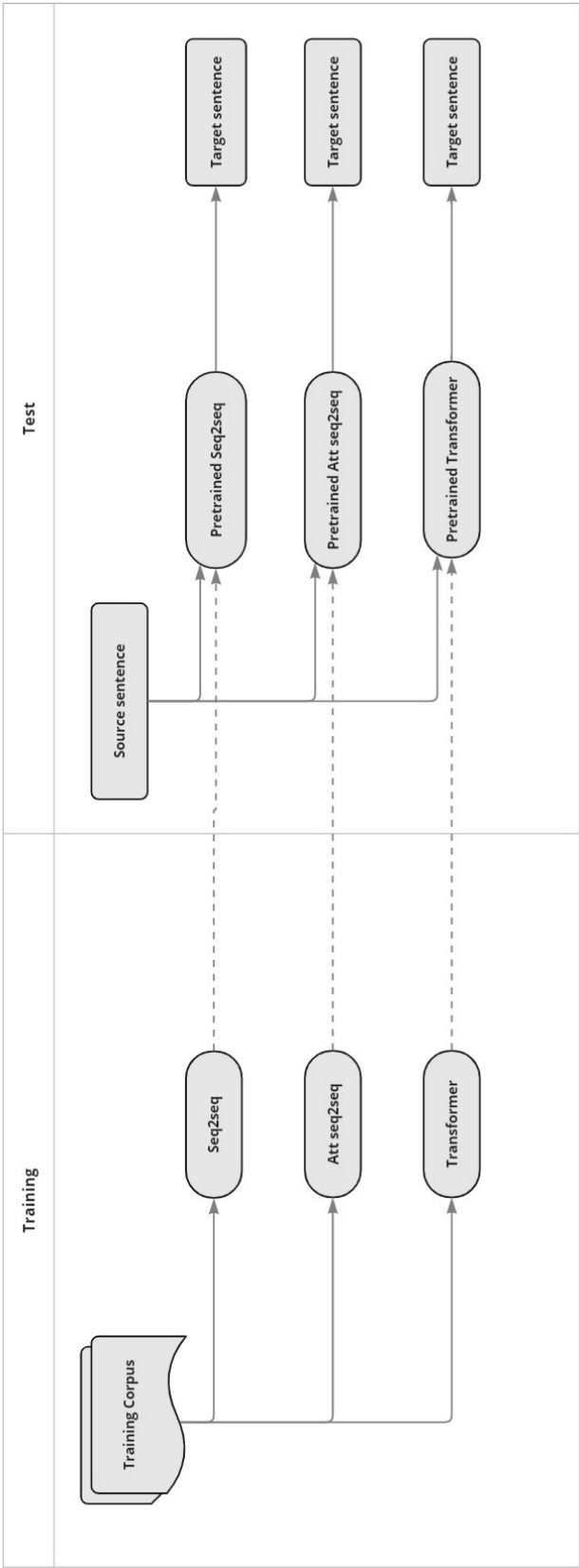


Figure 4.2: The NMT system’s architecture employed for translating Arabic dialects to MSA.

tures, vocabulary, and grammar, making it challenging for conventional machine translation models to accurately translate long sentences that carry all these linguistic complexities. Our approach leverages the advances in NMT and deep learning to produce high-quality translations that capture the unique linguistic and cultural features of the dialects being translated. We have also analyzed the impact of various model components and hyperparameters on the quality of the translations, and we have discussed the challenges and limitations of this approach.

4.4.1 Arabic Dialects Translation Using Seq2seq

The seq2seq translation model [79–81] consists of two main components: an encoder and a decoder as shown in Figure 4.3. The encoder is responsible for converting the source Arabic dialect sentence into a fixed-length representation, known as the encoding vector, which summarizes the information in the Arabic source sentence. The decoder, on the other hand, uses the encoding vector to generate the target MSA sentence. To handle the unique features of Arabic dialects, such as dialect-specific grammar and pronunciation, the encoder and decoder each consist of LSTM layers. These LSTM layers allow the seq2seq model to effectively capture the long-term dependencies between words in the Arabic dialect and MSA sentences, which is essential for generating accurate translations.

The encoder of a NMT model for Arabic dialects can benefit from stacking many LSTM layers. This added depth allows the network to better understand complex dependencies in the source Arabic Dialects sentence, resulting in more accurate translations. However, having too many layers can cause overfitting, reducing the model's performance on new data. Choosing the right number, type, size, and configuration of LSTM layers is crucial for optimal performance and requires careful tuning and experimentation. The seq2seq model in use has three LSTM layers in its encoder and one in its decoder.

Both encoder and decoder commence with an embedding layer. The embedding layer in NMT has a crucial impact on the performance of NMT in the context of Arabic dialect translation. The following points detail the ways in which the embedding layer can influence Arabic dialect NMT performance:

- The embedding layer is used to represent words as vectors in a high-dimensional space. This allows the NMT model to better capture the semantic and syntactic relationships between words, which can improve the overall performance of the model. For example, the word 'كوزينة' is used to denote "kitchen" in the Arabic dialect prevalent in Algeria. This term, however, is 'مطبخ' in MSA and cannot be identified by an MSA language model. The NMT model can better handle the representation of this dialect-specific term with a fine-tuned embedding layer, resulting in enhanced translation performance in the Algerian dialect.

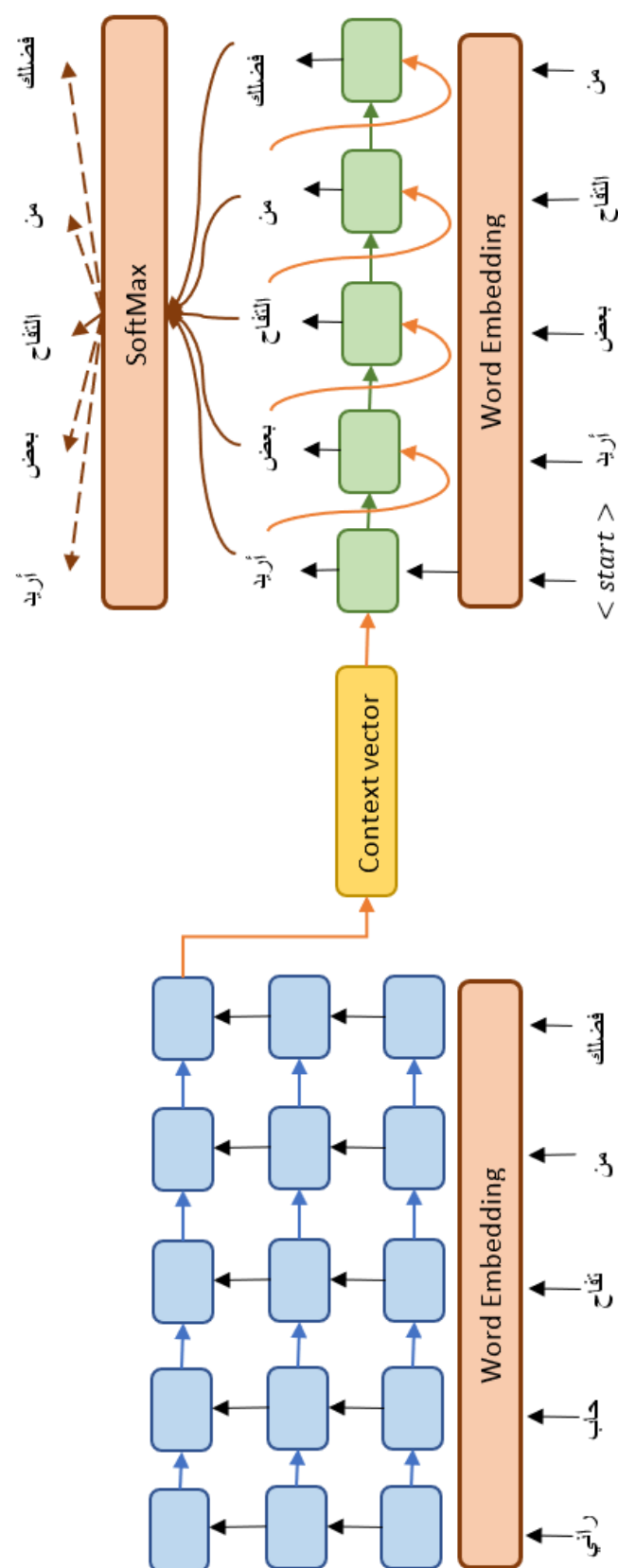


Figure 4.3: The Architecture of seq2seq NMT Model used to translate the Arabic Dialect.

- Several words in Arabic dialects may not be found in the MSA. The embedding layer enables the NMT model to handle these rare words and produce correct translations by capturing the semantic and syntactic meaning of words. For example, the words ‘براف’ and ‘برشا’ used in Algerian and Tunisian dialects, respectively, not be found in MSA, but they express the same semantic and syntactic meaning of the MSA word ‘كثيرا’.
- The embedding layer can be fine-tuned to handle out-of-vocabulary (OOV) words [82, 83] in the Arabic dialects, which can improve the overall performance of the model in translating dialect-specific words.
- Also, the embedding layer can be adjusted to better capture dialect-specific features such as pronunciation and grammar, leading to improved translation of dialect into MSA.

Ultimately, a SoftMax layer is situated at the top of the model. The SoftMax layer in a dialect-to-MSA NMT model generates a probability distribution for MSA vocabulary from dialect input. It calculates dot products using the decoder’s hidden state and weight vectors and then applies the SoftMax function to produce the distribution. The NMT model is trained to optimize this and produce the most likely MSA translation.

4.4.1.1 Experimentation

To evaluate the effectiveness of our seq2seq model on Arabic Dialect, we conducted several experiments using data sourced from the MADAR and PADIC multi-Arabic dialect parallel corpora. The process of dividing the MADAR and PADIC corpora into training and test data was a crucial aspect of our study.

As mentioned in Section 5.4, MADAR and PADIC are multidialectal parallel Arabic corpora with a wide range of Arabic dialects. MADAR corpus encompasses 25 distinct dialects, while PADIC corpus comprises 5 dialects. To effectively utilize these corpora, we meticulously divided the MADAR and PADIC corpora separately into separate components for training and testing purposes.

This division was carefully designed to ensure that the model could be trained on a sufficient amount of data and then tested on a representative sample to evaluate its performance. The division of the MADAR and PADIC corpora was specifically optimized to facilitate the training process and enable the model to learn the patterns and relationships between the data points effectively.

By dividing the corpora in this manner, we were able to achieve a well-trained model that could perform the desired task effectively. The experimental process is outlined in detail below.

Experiment 1 focused on assessing the model's ability to translate Algerian dialect using a dataset of 14026 sentences, consisting of PADIC sentences and additional ones related to COVID-19. During the training process, over 13000 sentences were utilized, with 10% of the dataset aside for validation and the remaining 893 sentences utilized for testing. The dataset was augmented multiple times to increase the number of sentences available for training [79].

Experiment 2 involved training the model on 10000 Algerian dialect sentences from the PADIC dataset and evaluating its performance on 2729 Algerian dialect sentences from the same dataset. During this experiment, it was noted that the Algerian dialect data comprised two distinct dialects, Annaba and Algiers, which led to the possibility of the test data including sentences from one dialect present in the training data but written in the other dialect [80].

In **Experiment 3**, the model was trained on 45000 sentences from the MADAR dataset, encompassing various Arabic dialects such as Aleppo, Alexandria, Algiers, Amman, Aswan, Baghdad, Basra, Beirut, Benghazi, Cairo, Damascus, Doha, Fes, Jeddah, Jerusalem, Khartoum, Mosul, Muscat, Rabat, Riyadh, Salt, and Sanaa. 10% of the data was used for validation, while 5000 sentences from Sfax, Tripoli, and Tunis Arabic dialects were employed for testing [81].

Experiment 4 utilized 34624 sentences from all Arabic dialects in the PADIC dataset to train the model, with 10% of the data reserved for validation and the remaining 3776 sentences utilized for testing the model's performance across all Arabic dialects in the PADIC corpus [81].

The experiments were designed to test the model's ability to translate multiple Arabic dialects, with each experiment utilizing different training and testing datasets. These experiments provide valuable insights into the model's effectiveness in translating Arabic dialects and may guide further research in this area.

The seq2seq model [79–81] was implemented using TensorFlow and Keras libraries. In order to evaluate the model's performance, various accuracy metrics and categorical cross-entropy loss were used.

To evaluate the performance of the model, the test data was utilized to compute the BLEU score value. The results were based on several experiments, each with a unique combination of batch size and epoch count. In the first experiment, the model was trained for a total of 30 epochs, with each epoch consisting of a selected batch size of 1024 sequences. The second experiment involved training the model for 50 epochs, with each epoch consisting of a selected batch size of 512 sequences. In the rest of the experiments, the model was trained for 50 epochs, with each epoch consisting of a selected batch size of 312 sequences. The selection of batch size and epoch count was done through a meticulous process of trial and error, with the primary objective being to achieve the best possible results. It's worth emphasizing that the selection of batch size and epoch count was not

arbitrary but based on rigorous experimentation.

In preparation for training, all sequences within a batch were padded to the length of the longest sequence in that batch. This was done to ensure consistency and to prevent any issues arising from varying sequence lengths.

4.4.1.2 Result

This section presents the findings of our experimental analysis. It includes a detailed explanation of the experimental results obtained, the score calculated using the BLEU score metric, and sample translations to demonstrate the performance of the model. This section aims to provide a comprehensive understanding of the effectiveness of the seq2seq model and its ability to generate translations.

Score BLEU Result

Experiment	Score Bleu
1	22.10
2	0.05
3	54.49
4	33.58

Table 4.2: BLEU scores results of four experiments evaluating the effectiveness of Seq2Seq model in Arabic Dialects translation.

The evaluation of the Seq2seq model's performance in the experiments relies on the BLEU score metric. Table 4.2 displays the BLEU score results for the Seq2seq model across four different experiments.

Overall, experiments 1, 3, and 4 show better results compared to experiment 2. The reason for this can be attributed to the size of the training data. In experiments 1, 3, and 4, we utilized larger training datasets, while, experiment 2 employed a comparatively smaller training dataset.

The larger size of the training dataset enables the models to more effectively capture the nuanced linguistic characteristics present in the dialects, resulting in the noticeable performance enhancements seen in experiments 1, 3, and 4. Nonetheless, it's essential to recognize that there are other factors to take into account, like the characteristics of the data, including syntax complexity, lexicon intricacy, and dialectal variations.

Furthermore, The seq2seq model demonstrate better results on the MADAR corpus (experiment 3) compared to the PADIC corpus (experiment 4). This improvement can be attributed to several factors:

- **Dataset size:** MADAR corpus is a larger corpus with more data, which allows the model to better capture the linguistic features of the dialects. This increased amount

of training data leads to a more accurate translation system.

- **Diversity of dialects:** MADAR corpus contains a more diverse set of dialects compared to PADIC corpus, which provides a wider range of linguistic features for the model to learn from. This leads to a more robust translation system that can handle a wider range of dialects.
- **Annotation scheme:** MADAR corpus has a more comprehensive annotation scheme, which includes morphological, syntactic, and semantic information. This additional information helps the model to better understand the linguistic structure of the dialects and to generate more accurate translations.
- **Sentence length:** MADAR corpus has shorter sentences compared to PADIC corpus, which can help the model to better focus on the local context of each sentence, leading to improved translation performance.

To summarize, the superior performance of the seq2seq model can be attributed to several factors. These include the larger size of the corpus, the greater dialectal diversity captured in the data, a more comprehensive annotation scheme, and shorter sentence lengths. These factors have led to an improvement in the accuracy and effectiveness of the translation models for Arabic dialects.

Translation Sample Result

In our study, we assessed the effectiveness of our NMT system for translating Arabic dialects using two corpora, MADAR and PADIC. To provide a clear representation of our findings, we have presented the results in tables, which are as follows:

- Tables 4.3, 4.5, 4.7, and 4.9 display the accurate translations generated by Seq2seq model for each experiment.
- Tables 4.4, 4.6, 4.8, and 4.10 showcase the partially correct translations produced by seq2seq model for each experiment.

The presented Tables provide compelling evidence of the NMT system's proficiency in accurately translating sentences of different lengths. While it is true that seq2seq model may occasionally generate incorrect translations for sentences of varying lengths, we aim to emphasize its overall strength in this regard.

In this study, sentence length was determined by the number of words, and the model performance was thoroughly evaluated on sentences of different lengths to assess its ability to handle both short and long sentences.

The results unequivocally demonstrate that the seq2seq model can effectively handle sentences of varying lengths and produce translations that are either accurate or partially accurate, depending on the length and complexity of the sentence.

Our system's accurate translations presented in Table 4.3 demonstrate the effectiveness of the seq2seq model in translating the Algerian dialect into MSA. The successful translation of these sentences is a testament to the model's ability to learn and generalize from the training data and capture the language's syntax, structure, and semantics.

However, it is essential to acknowledge that even with accurate translations, some variations or nuances may exist between the original Algerian dialect sentence and the MSA translation. Additionally, the performance of the model may vary based on the complexity and length of the sentences, the level of vocabulary used, and other factors.

Table 4.4 provides some valuable insights into the performance of the seq2seq model when it comes to translating the Algerian dialect into MSA. While the model was not able to translate all sentences accurately, it did manage to provide some partially correct translations for some of them.

Looking closely at the partially translated sentences, we can observe that the absence of certain particles or incorrect word translations had a significant impact on the accuracy of the translation. For example, in Sentence 1, the omission of the conjunction particle 'في' in the predicted translation resulted in an inaccurate translation that altered the intended meaning of the sentence. Similarly, in Sentence 2, the absence of the negation particle 'لا' in the predicted translation changed the context and meaning of the sentence.

Although the particle 'من' was replaced by 'و' in Sentence 3, the sentence's overall meaning was not affected. However, in Sentence 4, the incorrect translation of the word 'بعد' to 'لا' led to an inaccurate translation, even though the rest of the sentence was translated correctly.

For longer sentences such as Sentences 5 and 6, the model struggled to provide accurate translations, resulting in changes to the sentences' intended meaning. This is because some words present in the original translations were not present in the predicted translations, leading to altered contexts and meanings.

Overall, the results presented in Table 4.4 demonstrate the limitations of the seq2seq model when it comes to translating the Algerian dialect into MSA. In particular, the model's performance was hindered by a lack of sufficient training data, resulting in inaccurate translations in some instances. These findings highlight the need for more robust and diverse training data to enhance the model's accuracy and effectiveness in translation tasks.

According to the results presented in Table 4.5, the Seq2Seq model has shown remarkable efficiency in accurately translating short sentences, even with a relatively small blue score. This impressive success in producing correct translations for some sentences highlights the model's capability to handle short input sequences and indicates its proficiency in certain translation tasks.

Number	Size	Dialect	Original Translation	English	Predicted Translation	English
1	6	تعايبك واحد بنهمو صح تعرفوا لطاهر	سأعيايك شخصاً بنهمه حتا أعترف ال طاهر	I'll give you someone who truly understands, do you know Al-Taher?	سأعيايك شخصاً بنهمه حتا أعترف ال طاهر	I'll give you someone who truly understands, do you know Al-Taher?
2	7	مكاش كرونا في الجزائر كلش خراطي هذه مسرحية	لا يوجد كورونا في الجزائر كل شيء خرفات هذه مسرحية	There is no coronavirus in Algeria, everything is a lie, this is a play	لا يوجد كورونا في الجزائر كل شيء خرفات هذه مسرحية	There is no coronavirus in Algeria, everything is a lie, this is a play
3	8	العالم راه أعلن حالة طوارئ من شدة الخطورة	العالم بأسره أعلن حالة طوارئ من شدة الخطورة	The entire world has declared a state of emergency due to the severity of the danger	العالم بأسره أعلن حالة طوارئ من شدة الخطورة	The entire world has declared a state of emergency due to the severity of the danger
4	10	أنت لعزوز شفتي كيما تكون صابمة ديماء يصراولها الحوايج أوم	أفلم كلما كانت الوالدة صائمة تقع لها أمور كهذه	Do you know that whenever my mother fasts, things like this happen to her?	أفلم كلما كانت الوالدة صائمة تقع لها أمور كهذه	Do you know that whenever my mother fasts, things like this happen to her?
5	12	أوقفو اي مصدر للسكر السكر بين قص خمس خمسين بالمئة من جه از المناعة	أوقفو اي مصدر للسكر السكر يتقص خمسين بالمئة من جهاز المناعة	Stop any source of sugar, sugar reduces the immune system by fifty percent	أوقفو اي مصدر للسكر السكر يتقص خمسين بالمئة من جهاز المناعة	Stop any source of sugar, sugar reduces the immune system by fifty percent
6	21	من لديه المال يستطيع شراء مكلا ملاك غذائية عليكم بالسيلينيوم و الزنك و فيتامين دي3 و فيتامين سي و جولي روبايل	من لديه المال يستطيع شراء مكملات غذائية عليكم بالسيلينيوم و الزنك و فيتامين دي3 و فيتامين سي و جولي روبايل	Those who have money can buy nutritional supplements containing selenium, zinc, three types of vitamin D, vitamin C, and Royal Jelly	من لديه المال يستطيع شراء مكملات غذائية عليكم بالسيلينيوم و الزنك و فيتامين دي ثلاثة و فيتامين سي و جولي روبايل	Those who have money can buy nutritional supplements containing selenium, zinc, three types of vitamin D, vitamin C, and Royal Jelly

Table 4.3: Examples of Correct Translation Predictions by the Seq2seq Model in Experiment 1.

Number	Size	Dialect	Original Translation	English	Predicted Translation	English
1	11	حطو بيدكم على فمكم و ختمكم ك ي تكحو و لا تمكسو	وضع اليد على الفم والأنف في ح الة السعال أو العطس	covering the mouth and nose while coughing or sneezing	وضع اليد و الفم والأنف و حالة السعا ل أو العطس	The position of the hand, mouth and nose, and the case of coughing or sneezing
2	11	متدخلوش ليلايص فيها الغاشي لك ن كاين و لمان لازم متيقظ بأسر	اجتنب الدخول للأماكن المزدحم ة ان امكن وان كان الامر اضطر اريا لا يجب البقاء مطولا	Avoid entering crowded places if possible and if it is necessary, do not stay for a long time	اجتنب الدخول للأماكن المزدحمة ان امكن وان الامر اضطراريا و يجب البقاء مطول	Avoid entering crowded places if possible, even if the matter is urgent, and you must stay for a long time
3	12	لي يخدمو في الحوائت و الفارمسي لازم تهوية الدنيا حل البيبان و التوا في	لمن يعملون داخل المحلات التجا رية او الصيدليات يرجى توفير ا لتهوية اللازمة من فتح الابواب و التوافد	For those who work inside commercial stores or pharmacies, please provide necessary ventilation by opening doors and windows	لمن يعملون داخل المحلات التجارية و الصيدليات يرجى توفير التهوية اللا زمة و فتح الابواب والتوافد	For those who work inside shops or pharmacies, please provide the necessary ventilation and open doors and windows
4	15	بعد ما تسلم على الناس بليق متمس ش بيدك عينك و لا ختمك و لا هصك	بعد مصافحة الأشخاص يجب عد ف أو الفم	After shaking hands with people, it is advisable not to touch your hands to your eyes, nose, or mouth	لا مصافحة الأشخاص يجب عدم ملام سة الايدي العينين أو الأنف أو الفم	Do not shake hands with people. It is preferable not to touch hands with eyes, nose or mouth
5	24	اشرب تيزانة ناع الزنجبيل شخيرة و التوم ثانا خاطر ام يساهمو في ال قضاء على البكتيريا و يحو على ا لتكيف و ديرو سبور لكان كاين	استعمل الزنجبيل كمشروب ساخ ن وكذلك التوم لأنهما يساهمان في القضاء على البكتيريا والتهاد عن التدخين وممارسة الرياضة ا ن امكن	Using ginger as a hot drink and garlic because they help eliminate bacteria, avoiding smoking, and exercising if possible	استعمال الزنجبيل كمشروب ساخن و' كذلك التوم لأنهما يساهمان و القضاء و البكتيريا والتهاد و التدخين وممار سة الرياضة امكن	Using ginger as a hot drink, as well as garlic, because they contribute to the elimination of bacteria and avoid smoking and exercise as possible
6	40	عجبا لهذا الزمن تحزن ولا تجد من يواسيك تفرح ولا تجد من يهينك ت بكي ولا تجد من يمسح دمع عينك ولكن اعط فسوف تجد الكل ينظر إ ليك	عجبا لهذا الزمن تحزن ولا تجد م ن يواسيك تفرح ولا تجد من ير بك بككي ولا تجد من يمسح دم ع عينك ولكن اعطى فسوف تجد الكل ينظر إليك	Strange are the times we live in; you feel sad and find no one to comfort you, you feel happy and find no one to share your joy, you cry and find no one to wipe your tears, but if you make a mistake, you will find everyone watching you.	عجبا لهذا الزمن تحزن ولا تجد من يواسيك تفرح ولا تجد من يهينك تب كي ولا تجد و يمسح دمع عينك فسو ف اعطى فسوف تجد الكل ينظر إليك	It's strange in this day and age that when you're sad, you can't find anyone to comfort you, and when you're happy, you can't find anyone to share your joy. When you cry, there's no one to soothe you, for I will make a mistake, and you will find everyone looking at you

Table 4.4: Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 1.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	كيفاش	كيف ناك؟	How is that?	كيف ناك؟	How is that?
2	2	رائتي فونتريني	أنا أتمرر	I exercise	أنا أتمرر	I exercise
3	4	والله ما على بالي	صدقني لا أعلم	Believe me, I don't know.	صدقني لا أعلم	Believe me, I don't know.
4	4	ايه نكتبو أنا خير	نعم من الأحسن أن أكتبه بنفسي	Yes, it's better for me to write it myself.	نعم من الأحسن أن أكتبه بنفسي	Yes, it's better for me to write it myself.
5	6	خرجت حقيقت طلعت الدار خلاص ال نهال	ذهبت لحلق شعري وصعدت للمنزل إنتهي اليوم	I went to get my hair cut and then went back home, the day is over.	ذهبت لحلق شعري وصعدت للمنزل إنتهي اليوم	I went to get my hair cut and then went back home, the day is over.
6	12	لازم تمرروا بلي زوج لو تيرسلي عر ب جوا التوما و واحد من الأردن	لا بد أن تعلموا أن جامعتين عربيتين شاركنا أنتم و الآخر من الأردن	You must know that two Arab universities participated, one of them is yours and the other is from Jordan	لا بد أن تعلموا أن جامعتين عربيتين شاركنا أنتم و الآخر من الأردن	You must know that two Arab universities participated, one of them is yours and the other is from Jordan

Table 4.5: Examples of Correct Translation Predictions by the Seq2seq Model in Experiment 2.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	إليك الخط	إليك الخط	Here's the floor for you.	إليك عني	Here's the floor for you.
2	3	ما عذقتش مشكل	ما من مشكلة لديها	She doesn't have any problem	لا مشكلة في ذاك	There is no problem with that
3	3	ولاش راكي ديرى	ماذا تفعلين؟	What are you doing?	ماذا تفعلين به؟	What are you doing with it?
4	4	ما نعرف كيفاش تنكراتو	لا أعرف كيف تنكرته	I don't know how you remembered it	لا أعرف كيف عمله	I don't know how it works.
5	8	ما نعرف جددهم ويلا كان مزوج في عذلية	لا أعرف إن كان جددهم متزوجا في عذلية	I don't know if their grandfather was married in Anna ba.	لا أعرف إن كان جددهم متزوجا في ج امعة سيدي عاشور	I don't know if their grandfather was married at the University of Sidi Ashour.
6	11	صح ، لازم يتعاود كامل وتكون رقابة أكثر باش يخافوا شوية	صح ، يجب أن يعاد النظر فيه بأكمله وأن تكون هناك رقابة كي يخافوا قليلا	True, it should be completely reviewed and there should be supervision so that they fear a little.	صح ، يجب أن يعاد النظر فيه بأكمله وأن روى	True, it must be reconsidered in its entirety and insights

Table 4.6: Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 2.

Table 4.6 displays partially accurate translations achieved through the seq2seq model. As an illustration, consider sentence 1, where the original translation is ‘إليك الخط’ and the predicted translation is ‘إليك عني’. Both translations adhere to a basic clause structure and do not constitute complete standalone sentences. While the two sentences diverge in word choices, they partially overlap in meaning, implying that they convey related information to some extent.

Similarly, sentence 2, the original translation is ‘ما من مشكلة لديها’ and the predicted translation is ‘لا مشكلة في ذلك’. Although these translations differ in meaning, they are both negative sentences that emphasize the absence of the problem.

For sentence 3, the original and predicted translations have similar structure and wording. However, the addition of the preposition ‘به’ in the predicted sentence led to differences in meaning and context. The original translation is more general and open-ended, whereas the predicted translation is more focused and specific due to the inclusion of the preposition ‘به’.

Sentence 4, both original and predicted translations being negative declarative sentences. Although the first half of the predicted translation was translated correctly to ‘لا أعرف كيف’, the second half was translated incorrectly as ‘عمله’ instead of ‘تذكرته’.

In sentence 5, the original and predicted translations share a similar grammatical structure, with incorrectly translated particles in predicted translation. Lastly, sentence 6, the translation is only partially accurate, with only one-half of the sentence being correctly translated.

The results showcased in Table 4.7 highlight the exceptional ability of the NMT system to deliver precise translations of sentences with different lengths. The study reveals that the NMT system can effectively handle sentences of varying lengths and accurately translate them. Particularly, it is noteworthy that a significant number of correctly translated sentences comprise accurately translated sentences of different lengths, ranging from eight to eighteen words. These findings suggest that the NMT system has the capability to translate sentences with varying lengths accurately, regardless of the input sentence’s size, when using the MADAR corpus.

Table 4.8 displays partially correct translations generated by the NMT system. Upon inspection, it is evident that the predicted translation:

- For sentence 1, an error occurs in the predicted translation with the second word, where ‘ضاحية’ (oasis) is mistakenly translated as ‘بحيرة’ (lake). This inconsistency between the original and predicted translations implies that the model may encounter challenges in accurately translating specific Arabic words, especially those that are less frequent or possess multiple meanings.
- For sentence 2, the original and predicted translations are both in an interrogative form, but the sentence was translated incorrectly compared to the original translation

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	5	مشكل مكالمات يحكي ياباني، بمبتداك	أريد عامل تشغيل يتحدث اليابانية من فضلك	I need an employee who can speak Japanese, please.	أريد عامل تشغيل يتحدث اليابانية من فضلك	I need an employee who can speak Japanese, please.
2	8	مالا ايهبط الدروج هانم و ترة تشفوف تبعهم	بالتأكيد ، انزل من على الدرج هناك وسوف تجد بعض الإرشادات فالتبعها من فضلك	Sure, go down the stairs there and you will find some instructions, please follow them.	بالتأكيد ، انزل من على الدرج هناك وسوف تجد بعض الإرشادات فالتبعها من فضلك	Sure, go down the stairs there and you will find some instructions, please follow them.
3	8	سامحنى تتجمشى تقلى ولين الجم نلقى محطة تكسي؟	مخدرة هل لك ان تدلني على مكان موقف سيارات الأجرة ؟	Excuse me, could you please direct me to the location of the taxi stand?	مخدرة هل لك ان تدلني على مكان موقف سيارات الأجرة ؟	Excuse me, could you please direct me to the location of the taxi stand?
4	11	يوم الصحة والرياضه عترة أكتوبر عطلة رسميه على صحة الناس	يوم الصحة الرياضية هو يوم العاشر من أكتوبر وهو إجازة قومية الغرض منها تحسين صحة المواطنين	National Sports Day is on the 10th of October and it is a public holiday aimed at improving the health of citizens.	يوم الصحة الرياضية هو يوم العاشر من أكتوبر وهو إجازة قومية الغرض منها تحسين صحة المواطنين	National Sports Day is on the 10th of October and it is a public holiday aimed at improving the health of citizens.
5	18	اسمي غورو موتوياما التومرو هو صفر ثلاثة ثنين و ثلاثين سنين و اربعة و عشرين خمسة و تسعين	الاسم هو غورو موتوياما والرقم هو صفر ثلاثة ثلاثة اثنان سنة واحد اثنان أربعة تسعة خمسة	The name is Juro Motowayama and the number is zero three three two six one two four nine five	الاسم هو غورو موتوياما والرقم هو صفر ثلاثة ثلاثة اثنان سنة واحد اثنان أربعة تسعة خمسة	The name is Juro Motowayama and the number is zero three three two six one two four nine five
6	18	كان من تاخذ الرحلة رقم اثنين صفر، الي طلعت خمسطلن دقيقة فالت لكن تمام تقدر نظم رحلة اخرى	كان يفترض أن تسافر على الطائرة رقم ثمانية اثنان صفر التي أطلعت منذ خمس عشرة دقيقة حسنا بإمكانك تنظير رحلة أخرى	If you were supposed to travel on flight number eight two zero that took off fifteen minutes ago, well, I can arrange another trip.	كان يفترض أن تسافر على الطائرة رقم ثمانية اثنان صفر التي أطلعت منذ خمس عشرة دقيقة حسنا بإمكانك تنظير رحلة أخرى	If you were supposed to travel on flight number eight two zero that took off fifteen minutes ago, well, I can arrange another trip.

Table 4.7: Examples of Correct Translation Predictions by the Seq2seq Model in Experiment 3.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	وين الخواحي؟	أين الخواحي؟	Where is the suburb?	أين البيرة؟	Where is the lake?
2	4	تتجيم تحجز في المرض	هل يمكن أن تحجزوا في المرض؟	Can you make a reservation for the offer?	هل يمكن أن ترسل أحدا؟	Can you send someone?
3	4	وين أنجم نقاري كرهيتي	أين يمكن أن أوقف سيارتي؟	Where can I park my car?	أين يمكنني أن أستعيد أمتعتي؟	Where can I retrieve my luggage?
4	6	من فضلك كلمنا لما نكون واثي	سوف نتنظر من فضلك اتصل بنا عندما نكون مستعداً	We will wait, please contact us when you are ready.	من فضلك اتصل بي في الفندق عندما تجد حقائلي	Please call me at the hotel when you find my luggage
5	7	هو شويته غالي ما تعطيني خصم	إنه غالي الثمن إلى حد ما ألا يمكنك ملحي خصماً؟	It is quite expensive, can't you give me a discount?	إنه غالي جدا بالنسبة لي ألا يمكنك تخفيض السعر قليلاً؟	It's too expensive for me. Can't you lower the price a bit?
6	12	لازم نعطو الموضوع للتصويت؟ إذا لا، وصلنا إلى لمجلس إدارتنا غدوة المديح	هل نحتاج إلى طرح هذا للتصويت؟ إن لم تكن في حاجة إلى هذا، فإم كاني أن أعتقد حينئذ أننا قد وصلنا إلى نتيجة ساقوم بإرسالها إلى مجلس إدارتنا صباح غد	Do we need to put this to a vote? If we don't need to, then I guess we've reached a conclusion and I will send it to our board of directors tomorrow morning.	هل نحتاج إلى طرح هذا للتصويت؟ إن لم تكن في حاجة إلى هذا، فإمكانني أن أعتقد حينئذ أننا قد وصلنا إلى نتيجة ساقوم	Do we need to put this to a vote? If we don't need to, then I think we've reached a conclusion I will

Table 4.8: Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 3.

because only the two first words are translated correctly.

- sentences 3 and 4 are partially translated correctly, for sentence 3, only the first words are translated correctly, while for sentence 4, some words in the middle were translated correctly.
- For sentence 5, it was translated incorrectly compared to the original translation, but the overall meaning of the sentence is correct compared to the overall meaning of the original translation.

Table 4.9 provides empirical evidence of the NMT system's proficiency in accurately translating sentences of varying lengths. The outcomes demonstrate the NMT system's capability to handle sentences of different lengths and produce precise translations. The results indicate that the majority of the accurately translated sentences in the PADIC corpus comprise over 14 words.

Ultimately, Tables 4.9 and 4.10 presents compelling evidence of the NMT system's competence in translating sentences of different lengths with precision. The observations propose the system's ability to handle sentences of different lengths, including short and long ones, contingent on the corpus used.

Furthermore, Table 4.10 show the partially correct translations predicted by our NMT system, in which half of it was translated correctly and the rest was not. From the Table 4.10, we note that:

- Sentence 1 has been translated in a way that accurately conveys the intended meaning, despite the fact that the wording is different from the original. The word 'لا' is a negative particle that appears in both translations, indicating that the overall meaning of both sentences is negative. Furthermore, 'اعلم' and 'اعرف' are two different words in Arabic, but they both have the same meaning of "know". This demonstrates that even with a difference in wording, the intended meaning can still be accurately conveyed through translation.
- The predicted translation of sentence 2 is correct, but it is not a complete compared to of the original translation. The predicted translation is missing the word 'تماما' which indicates the near past. This missing information changes the context and the meaning of the sentence slightly, making it an incomplete representation of the original text. In translation, accuracy relies on correctly translating not just the words but also understanding and incorporating the context and nuances of the original sentence to convey the intended meaning effectively.
- In sentence 3, there is a significant disparity between the original sentence and the predicted translation. The translations have entirely different meanings. Addition-

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	14	عرفو عندي دوكيمون ديال الأمتال العربية، عزيز عليا الأمتال الحز الثرية و المغربية و التونسية	تلمون عندي مونة من الأمتال العربية أهوى الأمتال الحز الثرية و المغربية و التونسية	I have a blog of Arabic proverbs, and I particularly love Algerian, Moroccan, and Tunisian proverbs	تلمون عندي مونة من الأمتال العربية أهوى الأمتال الحز الثرية و المغربية و التونسية	I have a blog of Arabic proverbs, and I particularly love Algerian, Moroccan, and Tunisian proverbs
2	14	بعد المرات مالمين الأحوال الجوية، يتقولو بلي عندي تصيب ومن بعد السما صافية	أحيانا يخطي خبراء الأحوال الجوية يقولون بأنها ستمطر وبعد ذلك تكون السماء صافية	Sometimes weather experts make mistakes and say it will rain, but then the sky clears up	أحيانا يخطي خبراء الأحوال الجوية يقولون بأنها ستمطر وبعد ذلك تكون السماء صافية	Sometimes weather experts make mistakes and say it will rain, but then the sky clears up
3	15	ولكن هاد العام لقيت واحد الحاجة مكنتش وهي بلي مافيتش باش يمولوني فهاد الموضوع	لكن هذا العام وجدت شيئا لم أكن أنتظره وهو أنني لم أجد مؤطرا يوطرنني في هذا الموضوع	But this year, I found something unexpected which is that I didn't find anyone to support me in this matter	لكن هذا العام وجدت شيئا لم أكن أنتظره وهو أنني لم أجد مؤطرا يوطرنني في هذا الموضوع	But this year, I found something unexpected which is that I didn't find anyone to support me in this matter
4	16	مشي كبيرة تتهنن بلي بنتها الصغيرة في الشهادة ولا في الثانوية نخلات العام لي فات	ليست كبيرة أعتقد أن ابنتها الصغرى في قسم شهادة المتوسط أو في الثانوية نخلات العام الماضي	She is not old. I think her youngest daughter is in the Intermediate Certificate section or in high school. She entered last year	ليست كبيرة أعتقد أن ابنتها الصغرى في قسم شهادة المتوسط أو في الثانوية نخلات العام الماضي	She is not old. I think her youngest daughter is in the Intermediate Certificate section or in high school. She entered last year
5	20	مشينا جميع شوية من بعض رجعتا وباتت مريم هاف العرس هي تجيب لبنا المرض من تماك وجات	ذهبا جميعا جلسنا قليلا ثم عدنا وباتت مريم هنا في العرس فأحصرت ابنا المرض من هناك وجاءت	We all went and sat for a while, then we came back, and Mary stayed here at the wedding, so she brought us the disease from there and came	ذهبا جميعا جلسنا قليلا ثم عدنا وباتت مريم هنا في العرس فأحصرت ابنا المرض من هناك وجاءت	We all went and sat for a while, then we came back, and Mary stayed here at the wedding, so she brought us the disease from there and came
6	22	أنا باقي ا بابا و كيفاش بعيتي يكون خطي كيفاش بعيتي يطلع على راسي من السما و يخرج مابين صباغي ولا تنو	أنا أعلم يا أبي، وكيف تريد أن يكون خطي، ماذا تريد أن يسقط على الخط من السماء ثم يخرج من بين أصابعي أم ماذا؟	I am learning, Father. How do you want my handwriting to be? Do you want me to imagine a line falling from the sky and then writing between my fingers, or what do you want exactly?	أنا أعلم يا أبي، وكيف تريد أن يكون خطي، ماذا تريد أن يسقط على الخط من السماء ثم يخرج من بين أصابعي أم ماذا؟	I am learning, Father. How do you want my handwriting to be? Do you want me to imagine a line falling from the sky and then writing between my fingers, or what do you want exactly?

Table 4.9: Examples of Correct Translation Predictions by the Seq2seq Model in Experiment 4.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	مصر فتن	لا أعلم	I don't know	لا أعرف	I don't know
2	3	نعم تزوجت قبله	نعم تزوجت قبله تماما	Yes, I got married exactly before him.	نعم تزوجت قبله	Yes, I got married before him.
3	3	صدقني ولبت تقدر ترك	صدقني أتي صرت أشترط الآن	Believe me, I have become demanding now.	صدقني أنها أصحكتني كثيرا	Believe me, she made me laugh a lot.
4	4	أنا عفتها مرة واحدة	أحسنت عفتها مرة واحدة	You did well to contract it once.	و هل ساكلها مرة واحدة	Will I eat it once?
5	6	ظلت ليها عجلي لأما ولا بابا	ظلت ليها إتصلي بأما أو أبيها	I told her to call her mother or father.	ظلت ليها عليها أن تذهب و تجربها	I told her that she should go and try it.
6	25	تخيلي داك النهار كانت مع الزبون وهي من عاندتها تكلم تقول ليك جيلبي لي فهمتي أنا مع الزبون وهي تقول لي عجلي القلم	تخيلي في إحدى المرات كانت جالسة مع زبون وهي من عاندتها أن تكلم و لا تنهض تقول لك أحصري لي فهمت أنا مع الزبون و هي تقول لي أعطني القلم	Imagine one time she was sitting with a customer and it was her habit to give orders and no t move. She said to me, "Give me the pen" while I was with t he customer.	تخيلي في إحدى المرات كانت جالسة مع زبون وهي من عاندتها أن تكلم و لا تنهض تقول لك أحصري لي فهمت أنا مع الزبون	Imagine one time, a woman was sitting with a client and it was her habit to give orders and not get up. She said to you, "Bring me." Do you understand? She was with the client.

Table 4.10: Examples of Partially Correct Translation Predictions by the Seq2seq Model in Experiment 4.

ally, while the first word 'أين' meaning "where" was translated correctly, the remainder of the sentence was inaccurately translated.

- sentences 4 and 5 also was translated wrongly in comparison to the original translations and meaning, however the last two words from sentence 4 and the two first words from sentence 5 were correctly translated.
- The sixth sentence is partially translated correctly, but the translation is incomplete compared to the original translation.

Overall, these results suggest that the Seq2Seq model is proficient in handling short sentences but may encounter some difficulties when translating complex sentences.

4.4.2 Arabic Dialects Translation Using Attentional Seq2seq

The attentional seq2seq model [80, 81] depicted in Figure 4.4, is used to translate the Arabic dialect into MSA, their architecture is as follow:

- **Encoder:** The source sentence in the Arabic dialect is first passed through the encoder, which consists of an embedding layer and multiple LSTM layers. The embedding layer converts the input words into vectors, and the LSTM layers then compress the source sentence into a fixed-length representation.
- **Attention Mechanism:** The employed attention mechanism is bahundu attention. Additionally, the encoder produces a series of attention weights, corresponding to each word in the source sentence. These attention weights are used to dynamically focus on different parts of the source sentence when generating each word of the target sentence.
- **Decoder:** The decoder takes the fixed-length representation and the attention weights as input and generates the target sentence in MSA. The decoder also consists of an embedding layer and a single LSTM layer.

Embedding layers represent words in the Arabic dialects and MSA as continuous vectors to capture their semantic and syntactic relationships. The vectors are used as inputs for the LSTM layers, which can capture context and dependencies between words in a sequence, preserving the meaning and structure of sentences in the Arabic dialects. The attention mechanism uses this information to preserve essential details of the Arabic dialects while generating its corresponding MSA translation.

In the context of Arabic dialects translation, attention mechanisms in NMT can be used to give priority to certain words or phrases that are specific to the dialect, and to ensure that

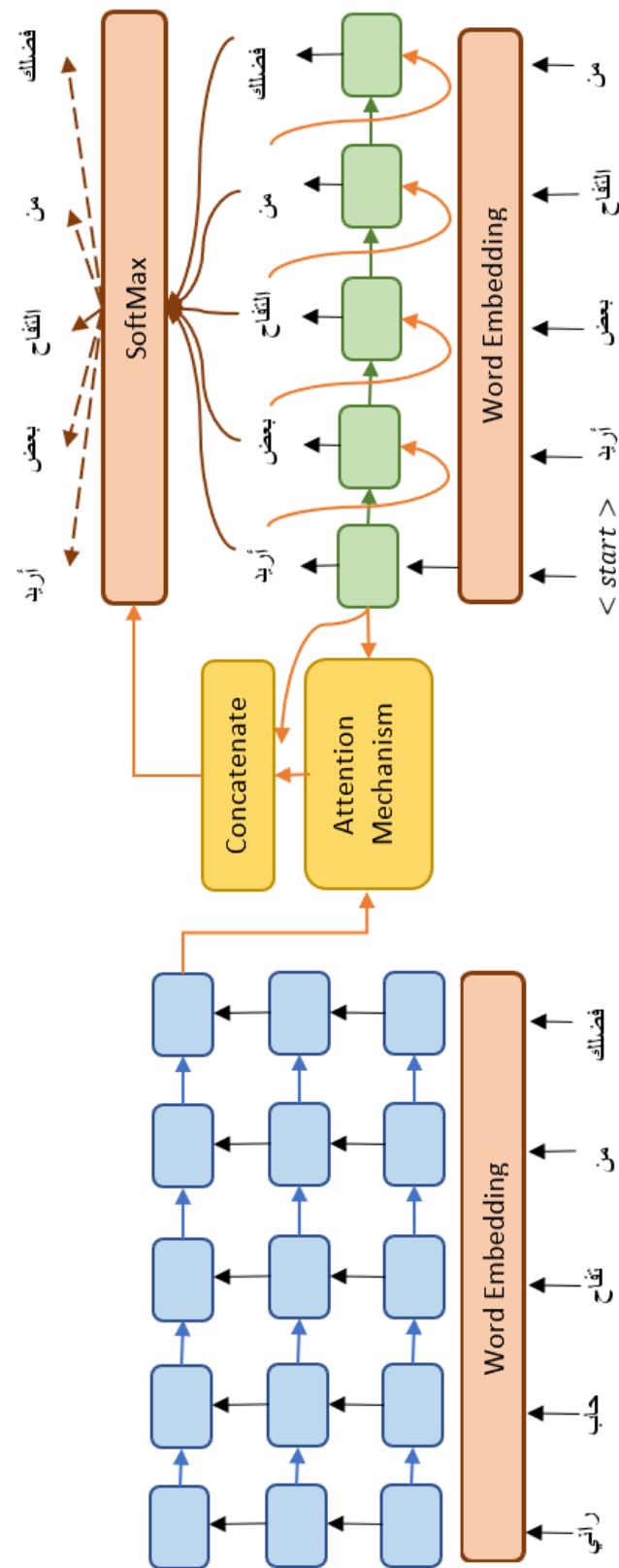


Figure 4.4: The Architecture of Attentional seq2seq NMT Model used to translate the Arabic Dialect.

these are accurately translated into MSA. By assigning higher attention weights to dialect-specific words or phrases, the NMT model can focus more on preserving the meaning of these words in the target language, which can help to produce more fluent and culturally-sensitive translations. Additionally, the attention mechanism can also help the NMT model to learn the relationships between different words in the dialect, which can further improve the quality of the translations.

Precisely, attention mechanisms can be leveraged in several key areas of the translation process:

- **Dialect-specific words:** If a word or phrase in the dialect has a distinct meaning compared to its equivalent in MSA, attention mechanisms can be used to give higher priority to preserving its meaning in the target language. For example, the attention mechanism prioritizes preserving the meaning of dialect slang technical words in MSA. also, it ensures the accurate translation of regional expressions specific to the dialect into MSA.
- **Dialect-specific grammar:** The grammar of a dialect may differ from MSA. Attention mechanisms can be applied to ensure the accurate translation of the dialect-specific grammar construct into MSA. For example, the word order in the dialect differs from MSA, the attention mechanism can be used to preserve the correct word order in MSA. Additionally, it can be used to ensure that unique tense forms, pronouns, and verb conjugations in the dialect are accurately translated into MSA. By preserving these linguistic details, the attention mechanism helps to maintain the overall coherence and fluency of the translated text.
- **Dialect-specific pronunciation:** Pronunciation in a dialect may differ from MSA. Attention mechanisms can be used to prioritize preserving these pronunciation differences in the target language, resulting in a more culturally sensitive translation. For example, the attention mechanism is key in keeping the phonetic, stress, and intonation details of a dialect intact during translation. It helps ensure that unique phonetic sounds, stress patterns, and intonation patterns of the dialect are accurately reflected in MSA.

Attention mechanisms provide a means of emphasizing critical elements of the dialect, such as dialect-specific vocabulary, syntax, and phonetics, in the translation process. This leads to a more authentic and culturally sensitive representation of the original dialect in the TL.

Finally, the output from the attention layer is combined with the output from the decoder and fed into the SoftMax, which generates the final translation output.

4.4.2.1 Experimentation

Our goal was to evaluate the effectiveness of the attentional seq2seq model on Arabic dialects. To achieve this, we conducted a series of comprehensive experiments using the multi-Arabic dialect parallel corpora provided in MADAR and PADIC corpora.

A crucial aspect of our study was the process of dividing the datasets into training and test data. We knew that the quality of our results depended on the accuracy and reliability of our data, so we meticulously analyzed and partitioned the datasets. This was done to ensure that our model was trained on a diverse range of examples, and to prevent overfitting, which can happen when the model are trained on too little data.

In addition to the experiments presented in Section 4.4.1.1, we conducted several other experiments to provide a more thorough understanding of the attentional seq2seq model's efficacy on Arabic dialects. These experiments included **Experiment 5**, where the model was trained on 11821 Algerian dialect sentences from both MADAR and PADIC corpora and evaluated on 3000 Algerian dialect sentences from the same corpora. We also conducted **Experiments 6 and 7**, where we trained the model on 7411 Tunisian and Moroccan dialect sentences, respectively, from the same corpora, and evaluated its performance on 3000 Tunisian and Moroccan dialects sentences from the same corpora.

To implement the model, we used TensorFlow and Keras libraries. In order to evaluate the model's performance, we used various accuracy metrics and categorical cross-entropy loss. We wanted to ensure that the model was performing optimally, so we carefully selected 512 batch size and 50 epochs count for each experiment. We used a process of trial and error to determine the best combination of batch size and epoch count, with the primary objective being to achieve the best possible results.

For all experiments, we carefully prepared the data by padding all sequences within a batch to the length of the longest sequence in that batch. This was done to ensure consistency and to prevent any issues arising from varying sequence lengths. After training, the results showed that all models were converging in terms of training, validation accuracy, and losses. This convergence indicates that the models were able to effectively learn the relationships and patterns within the data.

Overall, our experiments provide valuable insights into the model's effectiveness in translating Arabic dialects. The results can guide further research in this area and help to improve the accuracy of language translation models for Arabic dialects.

4.4.2.2 Result

In this section, we will present the experimental findings of our study, which will include a detailed explanation of the results obtained, the BLEU score calculated, and sample translations showcasing the performance of the attentional seq2seq model. The main goal of this section is to provide a comprehensive understanding of the model's effectiveness and

its ability to produce translations.

Score BLEU Result

Experiment	Attentional seq2seq
2	16.5
3	56.10
4	44.53
5	0.05
6	0.03
7	0.04

Table 4.11: BLEU scores results of seven experiments evaluating the effectiveness of Attentional Seq2Seq model in Arabic dialects translation.

The system evaluation employs the BLEU score metric, and Table 4.11 displays the results of the attentional seq2seq model across six experiments. The study's findings demonstrate that experiments 3 and 4 display superior model performance when compared to experiments 2, 5, 6, and 7. This contrast in performance might be attributed to the disparity in training data size, with experiments 3 and 4 benefiting from a larger training dataset. This larger dataset enables the model to better grasp the intricate linguistic nuances within the dialects, consequently leading to the observed performance boost in experiments 3 and 4. However, it's essential to acknowledge that additional factors, such as syntactic and lexical complexity, as well as the extent of dialectal variation, could also play a role in influencing the results.

Interestingly, the attentional seq2seq model's performance in experiment 2 surpassed that in experiment 5, even though they had similar training data sizes. This difference can be explained by the fact that the test data included sentences from one dialect, which were also present in the training data but written in another dialect. The similarity between these dialects likely contributed to the varying performance outcomes between the two experiments.

In conclusion, the enhanced performance of the attentional seq2seq model is attributable to several factors, including a larger dataset, greater dialectal diversity captured in the data, and a more comprehensive annotation scheme. Furthermore, the incorporation of the attention mechanism has significantly contributed to the development of a more accurate and robust translation system. Overall, these findings have led to an improvement in the accuracy and effectiveness of the translation models for Arabic dialects.

Translation Sample Result

In this part, we provide a comprehensive overview of the results obtained from the seq2seq model with attention mechanism, showcasing examples of correct, partially correct made

by the system. The results are presented in a tabular format.

Tables 4.12, 4.14, 4.16, 4.18, 4.20, and 4.22 display the correct predictions made by the NMT system on the experiments 2, 3, 4, 5, 6, 7, respectively. These tables showcase the instances where the system accurately translated.

In addition, Tables 4.13, 4.15, 4.17, 4.19, 4.21, and 4.23 present the partially correct results obtained in experiments 2, 3, 4, 5, 6, 7, respectively. These Tables display the instances where the system's predictions were partially correct, indicating that the system was able to capture some aspects of the translation.

In summary, this section objectively evaluates the performance of an NMT system in translating sentences of different lengths, highlighting its strengths and weaknesses. The Tables presented provide a clear representation of the system's performance, which facilitates a detailed analysis of its translation capabilities. It is evident from the results that the NMT system has a high proficiency in accurately translating sentences of different lengths. While there are occasional incorrect translations, the overall strength of the system in this regard is emphasized. The section aims to provide an evaluation of the NMT system's performance in translation.

Tables 4.12 and 4.13 show results from Experiment 2 with attentional seq2seq model. Table 4.12 presents the sentences that were translated completely correctly. We note that the model was able to translate sentences of varying lengths ranging from 2 to 21 words, and this highlights the impact of the attention mechanism.

Table 4.13 presents the sentences that were partially translated correctly. For sentence 1, the Algerian dialect sentence 'ماشي مشكل' is a common expression used to mean "no problem" or "it's okay". The NMT model's predicted translation, 'لا مشكلة في ذلك', is also an accurate translation, but it adds the phrase 'في ذلك' which means "in that" or "about that". This addition of 'في ذلك' is not necessary for the meaning of the sentence, but it does not change the overall message conveyed by the sentence.

For sentence 2, the Algerian dialect sentence 'نشا لله خير' was partially translated correctly by the NMT model. The predicted translation of 'إن شاء الله' is a valid translation, but it's not the complete translation of the original sentence. The NMT model seems to have missed the word 'خير' which means 'good', and thus it has provided a shorter translation.

For sentence 3, the Algerian dialect sentence 'عندي دوا تاع النسيان لازم تشريه', has the predicted translation 'هناك دواء للنسيان لن تتناوله' contains a negation, which contradicts the original sentence's declarative nature.

For sentence 4, the first half of the predicted translation, which is 'شحال راهي الساعة', was translated correctly by the NMT model. However, the second half of the sentence, which is 'على قداش بيداك الفيلم', was not accurately translated. The model instead provided a response that suggests a different context unrelated to the original sentence.

Referring to sentence 5, the second half of the predicted translation is relatively accurate. Nevertheless, the first half, 'أين ذهبت وأنت سأذهب أين', is incorrect, resulting in an inadequate representation of the overall meaning of the original sentence in the predicted translation.

For sentence 6, some words in the Algerian dialect sentence were correctly translated in the predicted sentence, while others were mistranslated or not translated at all.

Table 4.14 highlights the attention mechanism's effectiveness in NMT for translating longer sentences. However, our study demonstrates that the seq2seq model, when combined with the attention mechanism, can also achieve commendable performance with shorter sentences extracted from the MADAR corpus (experiment 3). Notably, our findings reveal that the model has accurately translated sentences of varying lengths, ranging from those with fewer than five words to over twenty-two words.

The results of our study suggest that the attentional seq2seq model has the potential to perform remarkably well across a broad spectrum of sentence lengths. This, in turn, provides greater flexibility and versatility in MT applications, enabling the system to handle diverse sentence structures and lengths, ultimately leading to more accurate and contextually appropriate translations.

In addition, upon closer inspection of the partially translated sentences from Table 4.15, we noticed that some were able to convey the intended meaning, despite not being translated word-for-word. For example, sentences 1 and 2 were not translated correctly, but their meanings were still conveyed accurately. The use of the negative particle 'لا' in sentence 1 indicated a negative form, and both 'اعلم' and 'أعرف' had the same meaning. In sentence 2, the phrase 'وهذا بشحال' was accurately translated to 'و هذا كم سعره' to inquire about the price. However, it should be noted that in MSA, 'كم سعره' or interrogative particles such as 'بكم' can also be used to ask about the price.

On the other hand, some of the partially translated sentences failed to capture the intended meaning. For instance, sentence 3 was partially translated correctly, with 'أه أنا' correctly translated to 'نعم أنا', but the second half 'قادرة' (able) was wrongly translated as 'رأيتها' (saw). Similarly, sentence 4 was partially translated incorrectly, with the latter half 'أنك لا يوجد من هذه' correctly translated to 'بلي حتى هدشي'. Sentences 5 and 6 also suffered from partial mistranslation.

Overall, the successful translation of a wide range of sentence lengths from the MADAR corpus (experiment 3) underlines the attention mechanism-equipped seq2seq model's immense potential in facilitating accurate translations across a variety of sentence structures and lengths.

Based on the findings presented in Table 4.16, it appears that the attentional seq2seq model, has the potential to perform well across a broad range of sentence lengths.

Upon analyzing the translations in Table 4.17, we have made several noteworthy ob-

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	هياك هي	وهو كذاك	And it is the same	وهو كذاك	And it is the same
2	5	هي راحت عند خالتي فاطمة	ذهبت عند خالتي فاطمة	I went to my aunt Fatima	ذهبت عند خالتي فاطمة	I went to my aunt Fatima
3	8	نشا الله يكون هذا الشيء في ميزان حسناتهم	أتمنى أن يكون هذا في ميزان حسناتها	I hope this will be in her good deeds' balance.	أتمنى أن يكون هذا في ميزان حسناتها	I hope this will be in her good deeds' balance.
4	21	نسنا نسنا باش يجولو على التسعة جيبنا ناع ناع نجمة و عيبنا لماما و عيبنا لي ناع بنت خال ناع باطل	انتظرت كثيرا ليتصلوا ، على التاسعة أخرجت شريحة نجمة الخا صمة بي و اتصالات بامي، فالتصالات بي من هاتف ابنة خالتي مجانا	I waited for a long time for them to call. At 9 o'clock, I removed my own Star SIM card and called my mom, but she called me back for free from my cousin's daughter's phone.	انتظرت كثيرا ليتصلوا ، على التاسعة أخرجت شريحة نجمة الخا صمة بي و اتصالات بامي، فالتصالات بي من هاتف ابنة خالتي مجانا	I waited for a long time for them to call. At 9 o'clock, I removed my own Star SIM card and called my mom, but she called me back for free from my cousin's daughter's phone.
5	15	قلت لهم اهدرو معاه و ضحكوه و اجبوا منه الهيرة ايدا قد كيما را ه يديوه ليسيكولوف	قلت لها تحدثوا معه و مازحوه و اسحبوا منه الكلام إذا بقي على ه ذا الحال خذوه إلى مختص نفساني	I told her to talk to him, joke with him, and get him to talk, and if he remains in this state, take him to a mental health specialist	قلت لها تحدثوا معه و مازحوه و اسحبوا منه الكلام إذا بقي على ه ذا الحال خذوه إلى مختص نفساني	I told her to talk to him, joke with him, and get him to talk, and if he remains in this state, take him to a mental health specialist
6	20	قادر يكون ثالث يوم قادر يكون رابع يوم راهم هدررو بالصبح الصبح ماراهوش باين دوك ماما راهي لثم دوك تنهم الصبح	إحتمال أن يكون ثالث يوم و احتم ال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أمني هنا ك ستقهم الوضع	It could be the third day or it could be the fourth day. They talked, but the truth is not clear. Mom, will understand the situation	إحتمال أن يكون ثالث يوم و احتم ال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أمني هنا ك ستقهم الوضع	It could be the third day or it could be the fourth day. They talked, but the truth is not clear. Mom, will understand the situation

Table 4.12: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 2.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	ماشقي مشكل	لا مشكلة	No problem	لا مشكلة في ذلك	No problem in that
2	3	نشأ الله خير	خيرا إن شاء الله	Good, God willing	إن شاء الله	God willing
3	6	عندي دوا ناع النسيان لازم تشربيه	لدي دواء للنسيان يجب أن تتناوليه	I have a medicine for forgetfulness that you should take.	هناك دواء للنسيان لن تتناوليه	There is a medicine for forgetfulness that you will not take.
4	8	شحال راهي الساعة على قد ش بيدا هداك الفيلم	كم الساعة؟ في أي ساعة يبدأ ذلك الفيلم؟	What time is it? At what time does that movie start?	كم الساعة؟ في الصباح وإن شاء الله	What's the time? In the morning, God willing
5	10	شوفي المكروب وبن كنت راي حة و وين جيت مت بالفرحة	إنه القدر، أين كنت سأذهب وأين أصبحت، أكاد أموت فرحا	It's destiny, where I was going and where I ended up, I could almost die of happiness	أين ذهبت وأنت سأذهب أين أنت أصبحت، أكاد أموت فرحا	Where did you go, and you, I will go, where have you become, I almost died of joy
6	19	راي نتعلم يا بابا كيفاش حيلنو الخط كيفاش الخط يطوح علينا من السما او مبد يخرج من صناعي ولا كيفاش	أنا أعلم يا أبي، وكيف تريد أن يكون خطي، ماذا تريد أن يسقط على الخط من السماء ثم يخرج من بين أصابعي أم ماذا؟	I am learning, my father. How do you want my handwriting to be? Do you want something to fall from the sky onto the paper and then to come out from between my fingers, or what?	أنا تمرقن يا أبي، وأنا تريد أن يكون خطي، ماذا تريد أن يسقط على الذي يسقط على الذي يخرج من بين أصابعي	I know, father, and I want my footsteps to be, what do you want to fall on me that falls on me that comes out through my fingers

Table 4.13: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 2.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	5	تخدم تحطلي في هيلتون، يعني مثلك؟	هلا أنزل أنتي عند فندق هيلتون من فضلك؟	Could you please drop me off at the Hilton Hotel?	هلا أنزل أنتي عند فندق هيلتون من فضلك؟	Can you drop me off at the Hilton Hotel, please?
2	6	شاك سياحي و بن ياباني، مر، يقل؟	هل يمكن استخدام تيكات سياحية و عملة بن ياباني؟	Can tourist checks and Japanese yen currency be used?	هل يمكن استخدام تيكات سياحية و عملة بن ياباني؟	Can tourist checks and Japanese yen currency be used?
3	11	وقت ترجع لليابان، باش نكتب جواب كان عندك وقت يمشك جواب	عندما أعود ثانية إلى اليابان، سأكتب خطاباً وإنا كان لديك وقت، من فضلك أرسل لي خطاباً	When I return to Japan again, I will write a letter, and if you have time, please send me a letter	عندما أعود ثانية إلى اليابان، سأكتب خطاباً وإنا كان لديك وقت، من فضلك أرسل لي خطاباً	When I return to Japan again, I will write a letter, and if you have time, please send me a letter
4	15	رمز ال منطقة هو صفر ثلاثة نومرو هو واحد اثنين ثلاثة أربعة خمسة ستة سبعة	كود المنطقة هو صفر ثلاثة والرقيم هو ثلاثة واحد اثنان ثلاثة أربعة خمسة ستة سبعة	The area code is zero-three and the number is one-two-three-four-five-six-seven.	كود المنطقة هو صفر ثلاثة والرقيم هو ثلاثة واحد اثنان ثلاثة أربعة خمسة ستة سبعة	The area code is zero three and the number is three one two three four five six seven.
5	17	بالحق؟ خايتي تبيت أوه، أنا متأسف برشا، تاكانو هادي عليلة من هادي فاتورتاك الجديدة، تخدم هو لي، يمشك؟	هل هذا صواب؟ دعني أتحقق أوه، أسف بشدة أفسه تاكانا هادي نا خطانا إليك فاتورتاك الجديدة هلا تخدمات بالتوقيع هنا، من فضلك؟	Is this correct? Let me check. Oh, I'm terribly sorry Miss Takanawa, this is our mistake. Here's your new invoice, would you please sign here, please?.....	هل هذا صواب؟ دعني أتحقق أوه، أسف بشدة أفسه تاكانا هادي فاتورتاك الجديدة هلا تخدمات بالتوقيع هنا، من فضلك؟	Is this correct? Let me check. Oh, I'm terribly sorry Miss Takanawa, this is our mistake. Here's your new invoice, would you please sign here, please?.....
6	20	قالي نحتلو بهنا لبار قال إني ماشي يعمل في تاليفون من غير ما يرجع جملة روجي تمام فاتورة عالية برشا	دعونا نحتفل، هكذا قال وأخذني لي إلى بار وقال إنه سيبري كلمة هاتية و غادر ولم يعد أبداً أو تخدم أنا وحدي عباءة فاقورة ضخمة	Let's celebrate, he said and took me to a bar. He said he would make a phone call and left, never to return. I was left alone to bear the burden of a huge bill.	دعونا نحتفل، هكذا قال وأخذني لي إلى بار وقال إنه سيبري كلمة هاتية و غادر ولم يعد أبداً أو تخدم أنا وحدي عباءة فاقورة ضخمة	Let's celebrate, he said and took me to a bar. He said he would make a phone call and left, never to return. I was left alone to bear the burden of a huge bill.

Table 4.14: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 3.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	3	نحب حديد حوايج	أريد مكواة	I want an iron	أريد مكواة سفر	I want a travel iron.
2	4	فشارتي مطعم رخيص هوني؟	هل يوجد مطعم أرخص قريب من هنا ؟	Is there a cheaper restaurant nearby?	هل يوجد مطعم تايلاندي قريب من هنا ؟	Is there a Thai restaurant nearby from here?
3	4	نحب نكلم سفارة اليابان	أريد أن اتصل هاتفياً بالسفارة اليابانية	I want to call the Japanese embassy by phone.	أريد الاتصال هاتفياً باليابان	I want to make a phone call to Japan
4	6	واحدة هامبرجر ووحدة بالجبن، من فضلك	واحد هامبرجر وواحد شيز بجر ، من فضلك	One hamburger and one cheeseburger, please	واحد شيز بجر لكي أذهب ، من فضلك	Can I have one cheeseburger to go, please?
5	8	الشعب الياباني تقريبا ميا وعشر مليون ساكن	تعداد سكان اليابان حوالي مائة وثمانين مليوناً	The population of Japan is about one hundred and twenty million.	إجمالي سكان اليابان حوالي مائة وثمانين مليوناً	The total population of Japan is around one hundred and thirty million
6	17	إيه أما تاتور عل ليسار في الخسوة الثالث وتبقى في هذا حتى الطريق يسكن وسكنين	نعم إنه هو ولكن سننصطر إلى الدوران يساراً في إشارة المرور الثالثة ونستمر في ذلك الشارع حتى نصل إلى الطريق رقم ستة وسكنين	Yes, it is it, but you will have to turn left at the third traffic signal and continue on that street until you reach Route Sixty-Six.	إنها بجور ذلك المبني المرتفع لذلك ارجع في نفس الطريق الذي أتيت منه حتى إشارة المرور ثم اتجه يساراً	It's next to that tall building so go back the same way you came to the traffic light and turn left

Table 4.15: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 3.

servations. Firstly, sentence 1 was translated accurately, but with the addition of the word 'سفر', which was not within the sentence's context and slightly altered its meaning. This highlights the importance of context in achieving accurate translations.

Secondly, although sentence 3 contains different wording from the original translation, it was translated correctly in terms of its intended meaning. This suggests that the NMT system was able to grasp the sentence's essence and effectively convey its message despite the variation in phrasing.

However, sentences 2, 4, 5, and 6 were translated partially incorrectly when compared to the original sentences, indicating that there is still room for improvement in the NMT system's ability to achieve precise translations.

Table 4.18 displays the accurate translations obtained from the experiment 5. The translations provided in the table represent the only correct translations obtained in the experiment, this is not surprising given the low blue score value obtained during the test phase. It is also worth noting that the model was only capable of translating short sentences with a maximum length of three words.

Table 4.19 contains partially correct translations from the same experiment. Sentence 1's translation is completely accurate and conveys the same meaning as the original sentence. However, the predicted translation includes the word 'نعم', meaning "yes", which, while not affecting the sentence's meaning, may reduce its accuracy according to the bleu score metric.

On the other hand, sentences 2, 3, and 4 have predicted translations that are entirely unrelated and do not make sense in the context of the Algerian dialect sentences. Although half of sentence 2, 'من فضلك', sentence 3, 'لماذا', and sentence 4, 'إنتظري قليلا', were translated correctly.

For sentence 5, some words are translated accurately, but the overall predicted translation fails to convey the Algerian dialect sentence's general meaning.

Regarding sentence 6, the predicted translation does not accurately represent the Algerian dialect sentence because it includes the name "فريدة", which is not part of the original translation. Additionally, the particle 'يا' is also absent in the original translation.

Table 4.20 presents the accurate translation results for Experiment 6, indicating that only two out of the six sentences were correctly translated. This outcome was expected due to the low blue score value obtained during the test phase.

In Table 4.21, partially correct translations from the same experiment were analyzed. The analysis revealed that for sentence 1, the first half of the sentence 'ذهبت إلى' was translated accurately, but the second half 'الجامعة' was incorrectly translated to 'نيويورك'. This translation error could be attributed to the model's limited understanding of the sentence's context and meaning, which is likely due to inadequate training data from the Tunisian dialect.

For sentence 2, while some words were correctly translated, the overall predicted trans-

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	5	باش ناخذ شهادة نخدم بها	حتى أحصل على شهادة أعمل بها	So that I can obtain a certificate to work with	حتى أحصل على شهادة أعمل بها	So that I can obtain a certificate to work with
2	8	فلش كملات قرأيتها قالت أيتها أنا ببع بت نخدم	لما أنهت دراستها قالت له أريد أن أعمل	When she finished her studies, she told him 'I want to work	لما أنهت دراستها قالت له أريد أن أعمل	When she finished her studies, she told him 'I want to work
5	13	الناس بدوا يهاد الموضوع حيث هو ما ف ميريكان بدوا تخدمو ف هاد الموضوع	الناس بدوا بالاهتمام بهذا الموضوع لأنهم في أمريكا بدوا العمل في هذا الموضوع	People started taking an interest in this topic because they began working on it in America	الناس بدوا بالاهتمام بهذا الموضوع لأنهم في أمريكا بدوا العمل في هذا الموضوع	People started taking an interest in this topic because they began working on it in America
3	15	إذا كنت طالبة عذري نيتشان اللي ما ستين بالمية تفتني والي مزوجة مية بالمية	إذا كنت طالبة فتمرين مباشرة لست متزوجة ستون بالمية تذهبين وأنت متزوجة مئة بالمية	If you are a student, you will practice directly. You are not married sixty percent, you go and you are one hundred percent married	إذا كنت طالبة فتمرين مباشرة لست متزوجة ستون بالمية تذهبين وأنت متزوجة مئة بالمية	If you are a student, you will practice directly. You are not married sixty percent, you go and you are one hundred percent married
4	19	يمكن تكون ثالث يوم و يمكن يكون رابع يوم راه هذرو ولكن الحقيقة ماما تماك عذري تقيم الوضع	احتمال أن يكون ثالث يوم و احتمال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أمة هناك ستقفم الوضع	There is a possibility that it could be the third day, and there is a possibility that it could be the fourth day. They have spoken to you, but the truth is not clear. Your mother is there, and she will understand the situation	احتمال أن يكون ثالث يوم و احتمال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أمة هناك ستقفم الوضع	There is a possibility that it could be the third day, and there is a possibility that it could be the fourth day. They have spoken to you, but the truth is not clear. Your mother is there, and she will understand the situation
6	22	لا لفرنسي و لا هم يحزنون حينئذ من راسك المربية قهنتي و غادي تهدي ب شكل عادي حيث اللغة ماشي بحال هكذا	لا الفرنسية لا تدرس و لا هم يحزنون، تنزعين من رأسك المربية قهنتي؟ و ستكلمين بشكل عادي لأن اللغة ليست هكذا	Not studying French is not a big deal, don't worry about it. Remove Arabic from your mind, do you understand me? You will speak normally because the language is not like that	لا الفرنسية لا تدرس و لا هم يحزنون، تنزعين من رأسك المربية قهنتي؟ و ستكلمين بشكل عادي لأن اللغة ليست هكذا	Not studying French is not a big deal, don't worry about it. Remove Arabic from your mind, do you understand me? You will speak normally because the language is not like that

Table 4.16: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 4.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	مصر فتنش	لا أعلم	I don't know	لا أعرف	I don't know
2	3	وهذا يشحال	و هذا بكم	And this is how much	و هذا كم سعره	And how much does it cost?
5	3	اه أنا قادة	نعم أنا قادرة	Yes I am able	نعم أنا رايتها	Yes, I saw her
3	5	قولي لي بلي حتى هدشي	قولي لي أنك لا تعلمين هذه أيضا	Tell me you don't know this either	قولي لي أنك لا يوجد من هذه	Tell me you don't have any of these
4	15	المره جايه فاش نقول ليك فيلم ولا مسلسل قولي لي تسناي كملتيه ولا لا	المره القادمه عندما أقول لك فيلم أو مسلسل قولي لي أكملته أم لا	Next time when I tell you a movie or series, tell me that you have completed it or not	ذات مرة قلت لك فيلم أو مسلسل ما كان أكملته أم لا	Once I told you whether a movie or series was completed or not
6	27	لكن عندك عائلة وقالت لك آ سوسن اجي انت معروضه عدا اذن اكيد غدي يوجد لك حاجه مزيانه من المسطح ومن بعد يوجد لك الحس باينه بك	لكن عندك قريب وقال لك يا سوسن تعالي انت مدعوة للخداء عدا إذا اكيد سيحضر لك شئيا جييا من المسطح ان بدعوك ثم يحضر لك الحس بديه يي أن يهتم بك	Next time when I tell you a movie or series, tell me that you have completed it or not	لكن عندك قريب وقال لك يا سوسن وسن تعالي انت مدعوة للخداء عدا إذا اكيد سيحضر لك شئيا جييا من المسطح ان بدعوك ثم يحضر لك الحس بديه	But you have a relative, and he said to you, Sawsan, come, you are invited to lunch tomorrow, so he is sure to bring you something good. It is impossible for him to invite you and then come.

Table 4.17: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 4.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	بياتسور	بالطبع	Of course	بالطبع	Of course
2	2	أه صحح	نعم صحح	Yes correct	نعم صحح	Yes correct
3	2	مش مشكل	لا مشكلة	No problem	لا مشكلة	No problem
4	2	ربي يحببناك	شكرا جزيل	Thank you very much	شكرا جزيل	Thank you very much
5	3	أيا باي باي	مع السلامة	Good-bye	مع السلامة	Good-bye
6	3	سباغا الحمد لله	بخير و الحمد لله	Fine and thank God	بخير و الحمد لله	Fine and thank God

Table 4.18: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 5.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	علائلي بيها	أنا أعلم ذلك	I know that	نعم أنا أعلم ذلك	Yes I know that
2	3	لطوكيو ، من فضلك	إلى طوكيو ، من فضلك	To Tokyo, please	سكين . من فضلك	Knife, please
3	3	وعلائش وانش معنتها	لماذا ما معني هذا؟	Why what does this mean?	لماذا ماذا يعمل	Why what works
4	4	أستاني تنوية راكي تهدي	انتظري قليلا، هل تتكلمين	Wait a minute, are you talking?	انتظري قليلا أظن أنني نجحت	Wait a minute, I think I succeeded
5	5	أو كاين دي سيبيسالياني بزاف	يوجد الكثير من الاختصاصات	There are many specializations	هناك الكثير من الدراسة	There is a lot of study
6	5	صباح الخير وانش راكي لايلس	أه صباح الخير ، كيف حالك؟	Ah good morning, how are you?	صباح الخير يا فريدة كيف الحال؟	Good morning Farida, how are you?

Table 4.19: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 5.

lation failed to convey the general meaning of the Tunisian dialect sentence. Similarly, for sentence 3, the first half of the sentence 'أين يمكنني' was translated accurately, but the second half 'أن أوقف سيارتي؟' was not a correct translation of 'تغير نقودي؟'. This can be attributed to the model's insufficient comprehension of the nuances and specificities of the Tunisian dialect.

In sentence 4, the model's predicted translation was grammatically incorrect, as it used the third-person masculine pronoun 'ه' instead of the second-person pronoun 'ك'. On the other hand, sentence 5 demonstrated that both the original and predicted translations were valid ways to say "Good morning" in different variations of Arabic, indicating that accurate translations may not always be word-for-word translations.

Finally, for sentence 6, while the model correctly included the greeting 'صباح الخير' and the question 'كيف حالك؟', it replaced the name 'ليبية' with the name 'فريدة' and altered the word order, indicating the model's limited understanding of the sentence's structure and content.

Number	size	Dialect	Original translation	English translation	Predicted translation	English translation
1	2	'أيه صحيح'	'نعم صحيح'	Yes, correct	'نعم صحيح'	Yes, correct
2	2	'يا ربي'	'يا إلهي'	Oh, my God	'يا إلهي'	Oh, my God

Table 4.20: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 6.

Table 4.23 presents the accurate translation results for Experiment 7, indicating that only five sentences were correctly translated. This outcome was expected due to the low blue score value obtained during the test phase.

Table 4.23 shows partially correct translations from the same experiment. The analysis revealed that for sentence 1, the model predicted the word 'المصعد' incorrectly to 'الإناء'. For sentence 2, in overall the predicted translation was correct but the addition of the name 'كمال' which is not exist in the Moroccan dialect sentence. For sentence 3, both the original and predicted translations convey the same general meaning, which is "Thank you, this is enough." However, there is a slight difference in the wording of the translations. The original translation reads 'شكراً، هذا يكفي' which includes the word 'يكفي' meaning "is sufficient" or "is enough". On the other hand, the predicted translation 'شكراً، هذا كافي' which uses the word 'كافي' meaning "enough" or "sufficient" without the auxiliary verb 'يكفي'.

For sentence 4, the predicted translation has two incorrect words 'بعض التفاح' instead 'منبه', changed the meaning of the Moroccan dialect sentence completely.

For sentence 5, the majority of words in the predicted sentence is incorrectly translated

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	مقديت لنيويورك	لقد ذهبت إلى نيويورك	I've been to New York	ذهبت إلى الجامعة	I went to the university
2	2	صباح النور	صباح النور	Good morning	صباح الخير	Good morning
3	3	قداش عمرها بنتك؟	كم عمر ابنتك؟	How old is your daughter?	كم عمر ابنته	How old is his daughter
4	3	بلاصة لنيويورك، وميتك	واحد لنيويورك ، من فضلك	One for New York, please	من فضلك أعطني دواء للبرد	Please give me medicine for a cold
5	4	ولين النجم نبدل فلووسي؟	أين يمكنني تغيير نقودي؟	Where can I change my money?	أين يمكنني أن أوقف سيارتي؟	Where can I park my car?
6	5	صباح الخير شحالك يا لبيبة	صباح الخير ، كيف حالك يا لبيبة	Good morning, how are you, Labiba?	صباح الخير يا فريدة كيف الحال؟	Good morning Farida, how are you?

Table 4.21: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 6.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	معلوم	بالطبع	Of course	بالطبع	Of course
2	2	صديقي تجرح	صديقي مصاب	My friend is injured	صديقي مصاب	My friend is injured
3	2	شكرا بالزاف	شكرا جزيل	Thank you very much	شكرا جزيل	Thank you very much
4	2	نعم بصح	نعم صحيح	Yes, correct	نعم صحيح	Yes, correct
5	2	يا ربي	يا الهي	Oh, my God	يا الهي	Oh, my God

Table 4.22: Examples of Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 7.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	3	فين هو المصعد؟	أين المصعد؟	Where is the elevator?	أين الإناء؟	Where is the pot?
2	4	صباح الخير كيف داير؟	صباح الخير كيف حالك	Good morning, how are you	صباح الخير كمال كيف حالك	Good morning Kamal, how are you
3	4	شكرا، هادشي كافي	شكراً، هذا يكفي	Thanks, that's enough	شكرا، هذا كافي	Thanks, that's enough
4	4	بغيت شوي منيه، عافاك	أريد منيه، من فضلك	I want an alarm, please	أريد بعض التفاح، من فضلك	I'd like some apples, please
5	6	الى خصه بيان، اتصل بيا عا فاك	إذا حدث ذلك، من فضلك اتصل بي	إذا حدث ذلك، من فضلك اتصل بي	أنكم، من فضلك ساو صالك إلى هاواي	Speak, please. I'll drive you to Hawaii
6	8	ممکن تحطيني جوج تذكيرات درجة ثانية لشيكافو، عافاك؟	هل يمكنك الحصول على تذكرتين درجة ثانية لشيكافو من فضلك؟	Can I get two second class tickets to Chicago, please?	هل يمكنك أن أطلع على شهادة التطعيم الخاصة بك من فضلك؟	May I see your vaccination certificate, please?

Table 4.23: Examples of Partially Correct Translation Predictions by the Attentional Seq2seq Model in Experiment 7.

leading to an incorrect translation.

For sentence 6, the majority of words in the predicted sentence is incorrectly translated leading to an incorrect translation.

To sum up, We noted that the attentional seq2seq model was able to construct translations with the correct context. These translations may be considered wrong compared to the original translation, but they have the correct meaning and convey the same intended message. This highlights the potential for different NMT models to improve the accuracy of MT and suggests that incorporating attention mechanisms may help improve the quality of MT.

Upon examining all presented tables, it is clear that the Attentional seq2seq model completely mistranslated some sentences of different lengths. The model's failure to generate appropriate translations could be attributed to its reliance on statistical patterns and lack of contextual understanding.

Upon conducting a comprehensive evaluation of our translation system, it has been observed that the attentional seq2seq model possess the capability to produce contextually appropriate sentences. Interestingly, even when the translation is considered incorrect, the results indicate that the NMT system can still convey the intended message and retain the original meaning. This is evident in examples 1 and 2 from Table 4.17, where the system generated an appropriate context despite the incorrect translation.

Furthermore, our NMT system has demonstrated the ability to predict the correct sentence form and type, including declarative forms such as sentences 3 and 6 in Table 4.17, and sentences 3, 4, and 5 in Table 4.23. It can also predict negative forms, as seen in sentence 1 of Table 4.17, and interrogative forms such as sentences 2 in Table 4.17 and sentence 1 in Table 4.23. This signifies that the system possesses a good understanding of the linguistic features of the dialects, including sentence structure and form.

In summary, the results of our evaluation demonstrate the NMT system's ability to produce contextually appropriate sentences while also predicting the correct sentence form and type.

4.4.3 Arabic Dialects Translation Using Transformer

We examined the efficacy of employing transformer model for translating Arabic dialects into MSA. Our primary goal was to assess the capabilities of the transformer NMT model in this specific task, considering its crucial role in NMT and substantial influence on MT research progress. We carried out our study with the main objective of evaluating the performance of this model in the complex task of translating Arabic dialects, which present distinct grammatical and lexical variations from MSA.

Attention-based seq2seq model employ an attention mechanism that enables the decoder to selectively concentrate on portions of the input sequence during the decoding

process (as discussed in the section 4.4.2). In contrast, transformer models utilize a self-attention mechanism that allows both the encoder and decoder to assess the significance of various parts of the input sequence while generating the output sequence. This self-attention mechanism is more potent than the attention mechanism used in attention-based seq2seq models, as it captures more intricate dependencies between words and produces more precise translations.

Moreover, attention-based seq2seq models typically rely on RNNs, such as LSTMs or GRUs, to encode and decode sequences. Meanwhile, transformer models exclusively use self-attention and feedforward layers. The transformer architecture does not incorporate RNNs, which enables it to parallelize training and inference more effectively than attention-based seq2seq models. This results in quicker training and inference times and makes transformer models better suited for large-scale MT tasks.

The transformer architecture employed is based on the same design outlined in section 3.4.3.5. By employing multi-head self-attention, the transformer model can simultaneously focus on various aspects of the input sequence, enabling it to effectively account for the subtleties in Arabic dialects' syntax and semantics.

Additionally, the transformer model benefits from the incorporation of positional encoding, which allows it to maintain the sequence's inherent order while processing input. This feature is particularly useful for handling the diverse word order variations and morphological structures present in Arabic dialects.

Another advantage of the transformer model in translating Arabic dialects into MSA is its layer normalization and residual connections, which improve gradient flow and mitigate the vanishing gradient problem.

Furthermore, the transformer model's flexibility in handling different input and output vocabulary sizes allows it to adapt to the diverse lexical inventory of Arabic dialects and MSA. This adaptability ensures that the model can effectively translate idiomatic expressions, regional vocabulary, and other language-specific nuances present in Arabic dialects.

4.4.3.1 Experimentation

Our research aimed to assess the effectiveness of the transformer model in translating Arabic dialects, using the multi-Arabic dialect parallel corpora provided in MADAR and PADIC. To achieve this goal, we conducted a series of comprehensive experiments using the same data division as experiments 5, 6, and 7 presented in Section 4.4.2.1.

We implemented the transformer model using the TensorFlow and Keras libraries and evaluated its performance using accuracy metrics and categorical cross-entropy loss. To ensure optimal performance, we carefully selected the batch size and epoch count for each experiment through a trial and error process. Our primary objective was to achieve the best possible results.

Table 4.24: BLEU scores results of three experiments evaluating the effectiveness of the Transformer model in Arabic dialects translation.

Experiment	Bleu score value
5	1.23
6	1.26
7	1.01

For all experiments, we prepared the data by padding all sequences within a batch to the length of the longest sequence in that batch. This step was taken to ensure consistency and prevent any issues arising from varying sequence lengths. After training, the results showed that all models converged in terms of training, validation accuracy, and losses, indicating that the models effectively learned the relationships and patterns within the data.

Our experiments provided valuable insights into the effectiveness of the transformer model in translating Arabic dialects. By fine-tuning its parameters on the multi-Arabic dialect parallel corpora, we demonstrated the potential of the transformer model in accurately capturing the nuances of Arabic dialects and producing high-quality translations.

4.4.3.2 Result

In this section, we will present the results of our experimental analysis, providing a thorough examination of our findings. This includes the evaluation score obtained from the BLEU metric and sample translations illustrating the Transformer model's performance. Our main goal is to offer a deep understanding of the model's effectiveness in generating translations.

Score BLEU Result

The system's performance was evaluated using the widely used BLEU score metric, which measures the quality of MT. Table 4.24 illustrates the results of the Transformer model from three distinct experiments, each of which involved a relatively small training corpus. The Algerian, Tunisian, and Moroccan dialects were trained on corpus consisting of 11,000, 7,000, and 7,000 sentences, respectively.

The results demonstrate that the Transformer model's performance was relatively low in all three experiments, which can be attributed to the limited size of the training corpus. These findings underscore the critical role of training data size in achieving better MT results.

Translation Samples Result

This section presents a detailed overview of the results obtained from the Transformer model, including both correct and partially correct translations. Tables 4.25, 4.27, and 4.29

showcase the correct predictions made by the NMT system, while Tables 4.26, 4.28, and 4.30 display the partially correct results.

The section provides an objective evaluation of the NMT system's performance in translating sentences of different lengths, highlighting its strengths and weaknesses. Despite occasional incorrect translations, the NMT system has a high proficiency in accurately translating sentences, as demonstrated by the tables presented. The aim of this section is to provide an unbiased evaluation of the NMT system's translation performance.

Table 4.25 presents the results of the transformer model's accurate translations, but the model's ability to translate was limited to very short sentences of three words or less. This limitation can be attributed to the small size of the training data used to train the model, which may not have provided enough exposure to diverse and complex language patterns.

However, Table 4.26 does expose some of the limitations of the model through its incorrect translations, it also showcases instances where the model has been successful in accurately translating certain linguistic contexts.

An illustration of this can be found in sentence 1, where the first word was translated accurately to 'كيف'. However, for the second word, the original translation was 'استطيع' whereas the model translated it as 'يمكنني'. Although the two words 'استطيع' and 'يمكنني' have the same context in this sentence.

Similarly, sentence 2, it is important to point out that the first half of the sentence was not translated accurately to 'مرحبا', which should have been translated as 'صباح الخير'. Nonetheless, both 'مرحبا' and 'صباح الخير' are commonly used as greeting words in Arabic. On the other hand, the second half of the sentence was translated correctly by the model.

Regarding sentence 3 had an interesting observation where the model correctly translated the first two words to 'لو كنت', despite the absence of 'لو' in the original translation. This suggests that the model may have been able to recognize a specific language pattern and apply it to the translation process.

Sentence 4 shows that the model struggled with sentences longer than three words, with only the first word being accurately translated. Likewise, in sentence 6, the entire sentence was translated incorrectly except for the word 'سأصل', which correctly conveyed the meaning of 'درك نعط'. These examples underscore the model's limitations in comprehending the subtleties of language and processing more complex sentence structures. The transformer model exhibits a noteworthy capacity for accurately translating short sentences, despite some instances of inaccuracy in the translations.

Table 4.27 displays the transformer model's precise translations, but it was only also able to handle very short sentences consisting of five words or fewer. This shortcoming can be also attributed to the small size of the Tunisian training data used to train the model.

Table 4.28 not only highlights the limitations of the model through its inaccurate translations but also demonstrates instances where the model has successfully translated specific

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	بياسور	بالطبع	Of course	بالطبع	Of course
2	1	ايه	نعم	Yes	نعم	Yes
3	2	باي باي	مع السلامة	Good bye	مع السلامة	Good bye
4	2	مساء الخير	مساء الخير	Good evening	مساء الخير	Good evening
5	3	ألو ايه فريال	ألو نعم فريال	Hello, yes Ferial	ألو نعم فريال	Hello, yes Ferial
6	3	مارا كيتس خايفة منو	لست خائفة منه	I'm not afraid of him	لست خائفة منه	I'm not afraid of him

Table 4.25: Examples of Correct Translation Predictions by the Transformer Model in Experiment 5.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	3	كيفاش تقدر نعلوك؟	كيف أسألتك مسألتك؟	How can I assist you?	كيف يمكنني أن أأخذ	How can i other
2	4	صباح الخير. واش راك؟	صباح الخير. كيف حالك	Good morning. How are you?	مرحبا كيف حالك؟	Good morning. How are you?
3	5	لو كان جيت في بلاصتك نشريها	كنت سأشتري هنا إن كنت في مكانك؟	I would buy this if I were in your place	لو كنت في الهاف التال	If you were on a mobile phone
4	5	سامحني، خالتي نخرج، من فضلك	عذراً. دعني أخرج، من فضلك	Excuse me. Let me leave, please	سامحني دعني منه من فضلك	Forgive me, please leave me alone
5	6	سأيا الحمد لله وثق راكي أنتيا	بخير و الحمد لله ،كيف حالك أنت؟	Fine, thank God, how are you?	بخير و الحمد لله الآخر و أنت بخير؟	Fine, the other, and you're fine?
6	10	درك نعط لها درك بعد و نعود نعط لـ فيقة نكو نقر ميلها	حسنًا سوف اتصل بها في الحال و أعلو د الاتصال برفيقة لأؤكد الأمر	Okay, I will call her immediately and I will call Ra feeqa again to confirm the matter.	الآن سألتك بها فيما بعد و أرى الميناء و الآخر	Now I'll call her later and see the port and the other

Table 4.26: Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 5.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	1	أيّه	نعم	Yes	نعم	Yes
2	1	يا لطيف	يا إلهي	Oh my God	يا إلهي	Oh my God
3	2	إنت هوني	أنت هنا	You're here	أنت هنا	You're here
4	2	دقيقة، رجعتك	لحظة من فضلك	Just a moment, please	لحظة من فضلك	Just a moment, please
5	5	أيّه أنا في الثانية ماستار	نعم أنا في الماستار الثاني	Yes, I'm in the second semester	نعم أنا في الماستار الثاني	Yes, I'm in the second semester
6	5	تقراو ثمنية سوايح في النهار	ندرس ثمانية ساعات في اليوم	We study for eight hours a day	ندرس ثمانية ساعات في اليوم	We study for eight hours a day

Table 4.27: Examples of Correct Translation Predictions by the Transformer Model in Experiment 6.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	3	شئوة نومرو الباب؟	ما رقم البوابة ؟	What is the gate number?	ما هو رقم الصراف ؟	What is the automated teller machine number?
2	3	هنا كل شيء؟	هل هنا كل شيء ؟	Is that everything?	هنا كل ما كان	That's all there was.
3	3	بحب كعكة، يميّتك	أريد كعكة محلاة من فضلك	I would like a sweet cake, please.	أريد أن أسألكم من فضلك	I want to ask you, please.
4	3	بحب شوية قهوة؟	ما رأيك في بعض القهوة ؟	What do you think about having some coffee?	هل تريدون أن أخذ بعض القهوة ؟	Do you want me to get some coffee?
5	4	وين تلقا دواء الرأس؟	أين يمكنني الحصول على أسبرين ؟	Where can I get some aspirin?	أين يمكنني أن أتناول الدواء ؟	Where can I have lunch?
6	8	وربلي الحوت الى ربتنه اليوم في الكتاب، يميّتك	أرني السمكة التي رأيته اليوم في كتاب الصور ، من فضلك	Please show me the fish you saw today in the photo book.	أرني بعض الأشياء التي اليوم في البداية ، من فضلك	Please show me some things that were there at the beginning of today

Table 4.28: Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 6.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	2	فين خدام	أين يعمل؟	Where does he work?	أين يعمل؟	Where does he work?
2	3	اه هو مزيان	نعم إنه جيد	Yes, he is good.	نعم إنه جيد	Yes, he is good.
3	3	الله يكون ف عونكم	ليكن الله في عونكم	May God assist you.	ليكن الله في عونكم	May God assist you.
4	4	مزيان إن شاء الله	حسنًا، إن شاء الله	Okay, God willing.	حسنًا، إن شاء الله	Okay, God willing.
5	4	شكرا سارة شكرا برف	شكرا سارة، شكرا جزيلا	Thank you, Sarah. Thank you very much.	شكرا سارة، شكرا جزيلا	Thank you, Sarah. Thank you very much.
6	6	شكرو قالت ليك شكرو قالت ليك	ماذا قالت لك؟ ماذا قالت لك؟	What did she say to you? What did she tell you?	ماذا قالت لك؟ ماذا قالت لك؟	What did she say to you? What did she tell you?

Table 4.29: Examples of Correct Translation Predictions by the Transformer Model in Experiment 7.

Number	size	dialect	Original translation	English translation	Predicted translation	English translation
1	3	شكراً ، هادشكي كافى	شكراً ، هذا يكفي	Thank you, that's enough.	شكراً ، هذا كافى	Thank you, that's enough.
2	3	بيعت نكري بار اسول	أود استئجار شمسية	I would like to rent a solar panel, please.	أريد استئجار كرسي مركب	I want to rent a reclining chair.
3	4	صباح الخير. كيف داير؟	صباح الخير . كيف حالك	Good morning, how are you?	صباح الخير كيف حالك يا أمي كم الأجرة ؟	Good morning, how are you, Mom? How much is the rent?
4	4	بيعت شي منليه، عفاك	أريد منليه ، من فضلك	I want an alarm clock, please.	أريد بعض التفاح ، من فضلك	I would like some apples, please.
5	5	عفاك نظف ليا هاد الكوط	من فضلك نظف هذا المعطف	Please clean this coat.	من فضلك أظهرها هذا المعطف	Please show me this coat.
6	13	يوم الصحة والرياضة عشرة أكتو بر هو عطلة وطنية كنهف لتعزير الصحة دبال الشعب	يوم الصحة الرياضية هو يوم العا شتر من أكتوبر وهو إجازة قومية ال غرض منها تحسين صحة المواطن ين	Health and Sports Day is on October 10th and it is a national holiday aimed at improving the health of citizens.	يوم الصحة والرياضة في أكتوبر في أكتوبر هو الملف الذي يحتله لكم أمس الثلاثاء عندما يوم صباحا	Health and Sports Day in October in October is the file I sent you yesterday, Tuesday. We have a morning on that day.

Table 4.30: Examples of Partially Correct Translation Predictions by the Transformer Model in Experiment 7.

linguistic contexts. For instance, in sentences 1, 2, 3, and 6, some words were translated accurately, but other words were not.

Regarding sentence 4, the predicted translation is not an accurate translation. While it is a question about whether the person wants the speaker to take some coffee, rather than asking if the person likes or wants some coffee. Overall, the predicted translation by the model is not accurate and fails to convey the intended meaning of the dialect sentence.

An interesting observation regarding sentence 5 is that while the second half was not translated correctly, the model did use the appropriate verb with the object, indicating an understanding of the context of the sentence despite the inaccurate translation.

Table 4.29 exhibits the accurate translations produced by the transformer model, but it also had limited capability in handling sentences with only six words or less. This limitation can be attributed to the inadequacy of the Moroccan training data used to train the model.

In contrast, Table 4.30 not only showcases the model's limitations by displaying its inaccurate translations, but it also illustrates specific linguistic contexts where the model was able to produce successful translations. For instance, in Sentence 1, the model performed exceptionally well in producing an accurate translation. However, it replaced the word 'يكفي' with 'كافي', which has the same meaning.

Additionally, while many words were translated correctly, some were substituted with others that carried different meanings. For example, in Sentence 4, the word 'منبه' was replaced with 'التفاح', and in Sentence 5, the verb 'نظف' was replaced with 'أظهر', resulting in a significant change in meaning. In Sentence 6, several words were repeated, leading to an incorrect context.

4.4.4 Discussion

One of the key challenges in MT is to effectively translate Arabic dialects, particularly the Maghreb dialect, which requires various techniques to make them understandable to everyone. To address this, three models of seq2seq, attentional seq2seq and transformer are used to translate Arabic dialects.

In this section, we explore the NMT models for Arabic dialects and investigate the impact of attention and data size on improving NMT models. To guide our discussion, we ask several questions. Firstly, we examine whether the Arabic dialects NMT model performs better with or without the attention mechanism, to determine the impact of attention on the NMT system in general. Secondly, we investigate the optimal sentence length for MT, to identify if there are specific sentence lengths that yield better NMT results. Lastly, we explore the crucial question of the impact of data size of Arabic dialects on the NMT system.

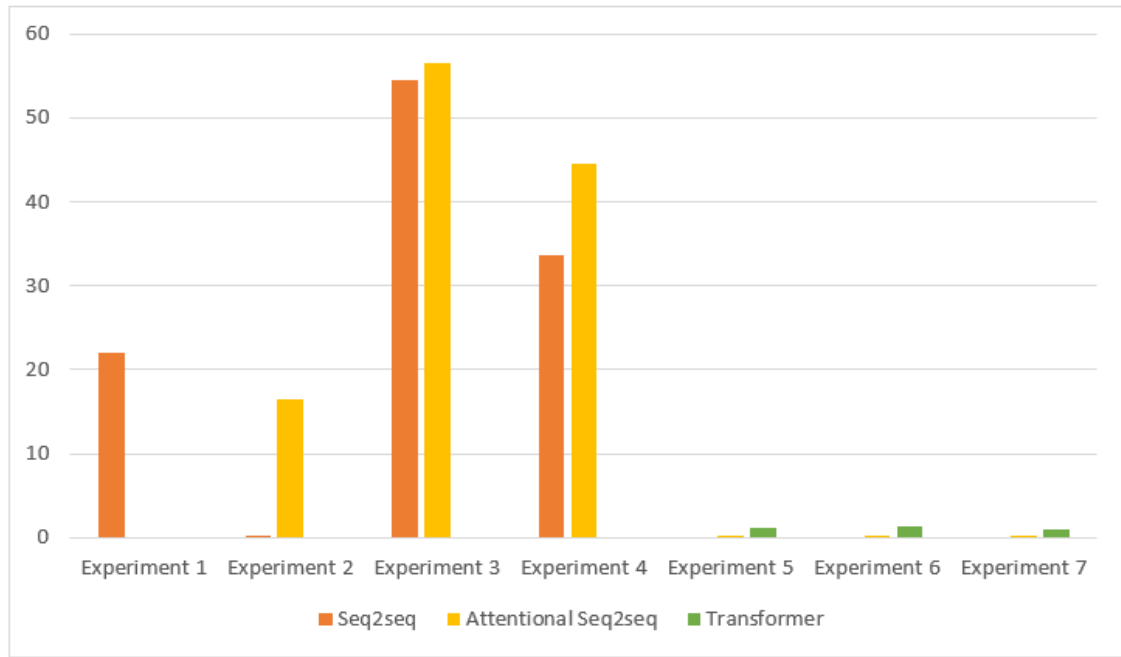


Figure 4.5: Comparison of BLEU score results for each experiment with different models: Seq2Seq, Attentional Seq2Seq, and Transformer. The figure showcases the performance of each model in Arabic dialect NMT tasks, highlighting their effectiveness in terms of BLEU scores.

Do Arabic Dialects NMT Model Performed Better With or Without Attention Mechanism?

In our previous experiments, we utilized three different models to translate Arabic dialects: seq2seq (sections 4.4.1), Attentional seq2seq (section 4.4.2), and Transformer (section 4.4.3).

After conducting the test phase, we obtained the blue score results, which are as follows (Figure 4.5 gives a better presentation to the obtained results):

- With seq2seq Model:
 - For experiment 1, it yielded a score of 22.10.
 - For experiment 2, it yielded a score of 0.05.
 - For experiment 3, it yielded a score of 54.49.
 - For experiment 3, it yielded a score of 54.49.
- With attentional seq2seq Model:
 - For experiment 2, it yielded a score of 16.5.
 - For experiment 3, it yielded a score of 56.10.
 - For experiment 4, it yielded a score of 44.53.

- For experiment 5, it yielded a score of 0.05.
- For experiment 6, it yielded a score of 0.03.
- For experiment 7, it yielded a score of 0.04.
- With Transformer Model:
 - For experiment 5, it yielded a score of 1.23.
 - For experiment 6, it yielded a score of 1.26.
 - For experiment 7, it yielded a score of 1.01.

In experiment 2, the seq2seq model yielded a very low score of 0.05, indicating poor performance in translating Arabic dialects. However, the attentional seq2seq model performed much better with a score of 16.5, indicating the attention mechanism's effectiveness in improving translation accuracy.

In experiments 3 and 4, the attentional seq2seq model outperformed the seq2seq model, with scores of 56.10 and 44.53, respectively, compared to scores of 54.49 and 33.58 for the seq2seq model. These results suggest that the attention mechanism improves the performance of the seq2seq model in Arabic dialect translation.

The results of experiments 5, 6, and 7 indicate that the transformer model with attention in both its encoder and decoder outperformed the attentional seq2seq model, achieving a 1 point higher bleu score for Algerian, Tunisian, and Moroccan dialects.

Based on our analysis, we found that the attentional seq2seq model outperformed the seq2seq model on both corpora. Specifically, in experiment 4 with PADIC corpus, the attentional seq2seq yielded superior results compared to the seq2seq model, and the same held true for experiment 3 with MADAR corpus. These findings provide compelling evidence that the attention mechanism significantly enhances the performance of NMT. Additionally, the transformer model, which is essentially based on the attention mechanism, outperformed the attentional seq2seq model.

Overall, our results demonstrate the critical role of the attention mechanism in improving the accuracy of NMT for Arabic dialects.

At What Sentence Length Can NMT Work Best?

Figure 4.6 showcases a detailed analysis of the BLEU score for sentences segmented by their length for each experiment. The results are presented in figure 4.6, and we derived some interesting observations from the data.

Upon analyzing the figure, it was determined that the seq2seq, attentional seq2seq, and Transformer models were most effective for sentences of moderate length. Specifically, the highest BLEU score was attained under the following conditions:

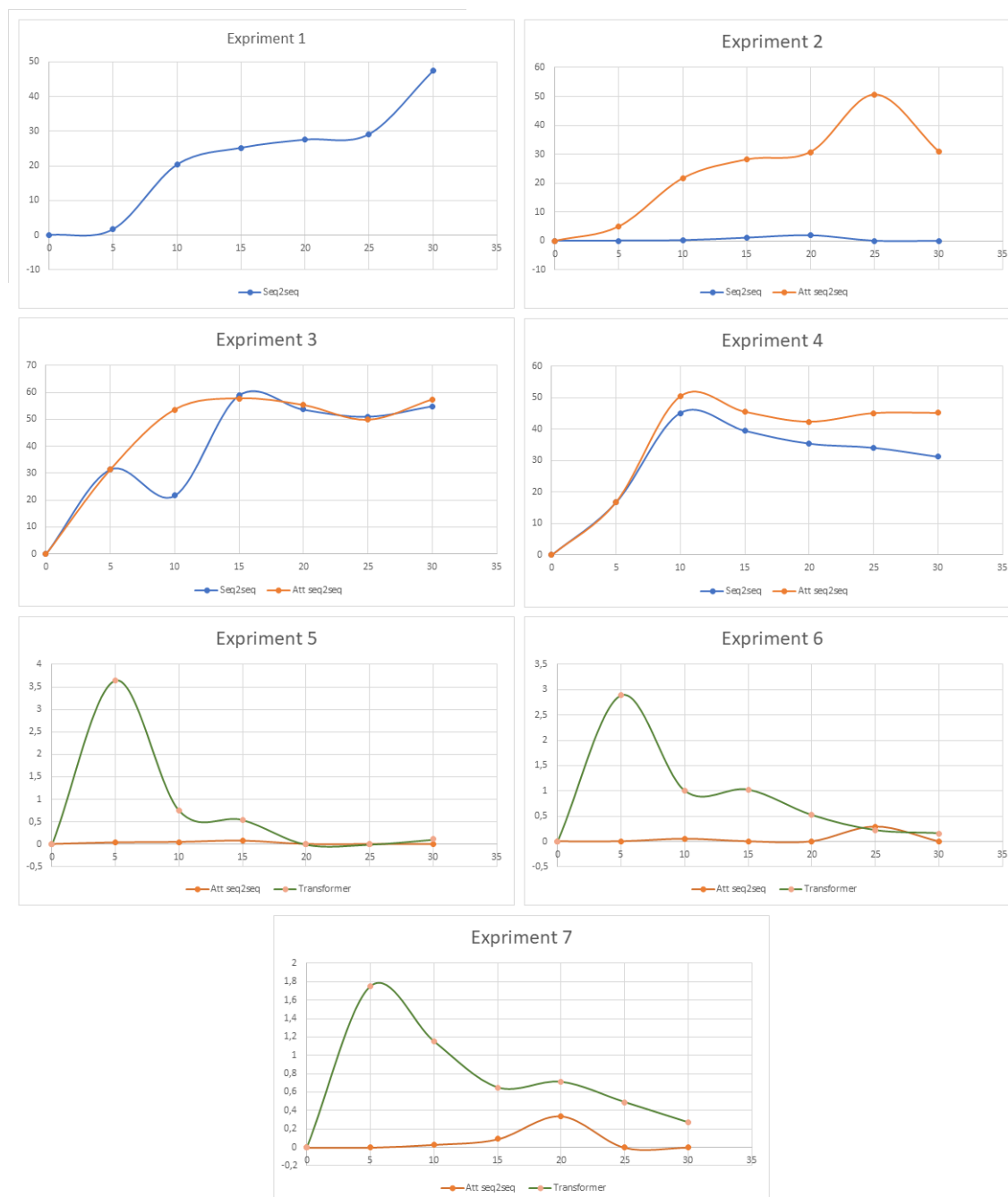


Figure 4.6: BLEU score value relative to the length of the sentences for each experiment.

- Experiment 1: For sentences containing more than 25 words
- Experiment 2: For sentences containing 15 to 25 words
- Experiment 3: For sentences containing 10 to more than 25 words
- Experiment 4: For sentences containing 5 to 15 words
- Experiment 5: For sentences containing 0 to 15 words
- Experiment 6: For sentences containing 0 to 15 words with transformer and 20 to 25 words with attentional seq2seq.
- Experiment 7: For sentences containing 0 to 15 words with transformer and 15 to 20 words with attentional seq2seq.

This is because these sentences are easier to comprehend and contextualize, resulting in mostly accurate translations. Although the longest sentence lengths appeared to work best for the Algerian dialect, the results cannot be considered reliable due to the significant variation in the number of sentences for each length.

Upon comparing the seq2seq, attentional seq2seq, and Transformer models, we discovered that the attentional seq2seq consistently outperformed the seq2seq model. Furthermore, the Transformer model surpassed the performance of the attentional seq2seq model in all groups. This can be attributed to the attention mechanism, which allows the model to concentrate on specific segments of the input sequence while producing the output sequence. By doing so, the attention mechanism helps to overcome the issue of "vanishing gradient" that is frequently observed in the seq2seq model.

In summary, the analysis of the BLEU score by sentence length provides crucial insights into the performance of the seq2seq, attentional seq2seq, and Transformer models. The findings suggest that the models perform best for medium-length sentences.

Comparing our Work With Other Research

Figure 4.7 provides a comparison between our research and other studies that have delved into Arabic dialects translation. Its purpose is to evaluate the efficacy of our work in comparison to other contributions in this domain. After carefully scrutinizing the data presented in the Figure, it became clear that the translation systems we utilized outperformed those of other studies by a significant margin, demonstrating the effectiveness of our experiments.

The Impact of Data Size on The NMT System

Utilizing deep neural networks, NMT is a method that learns the mapping between a SL and a TL. A significant challenge in constructing an effective NMT system is the issue of data size.

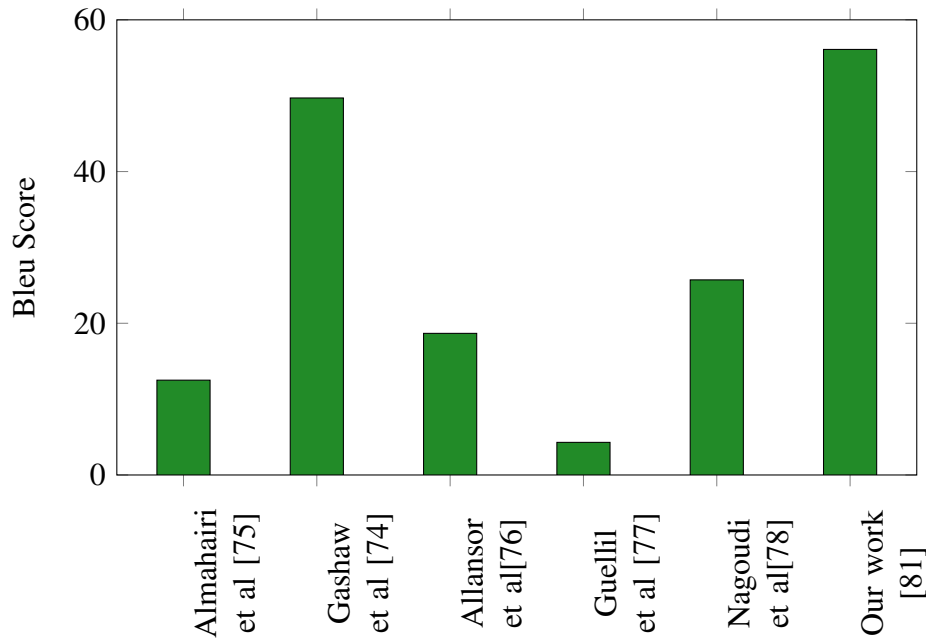


Figure 4.7: Comparison of Results from our Work and Related Studies.

Upon analyzing the results of our experiments, we observed that our system produced satisfactory outcomes in some cases, such as experiments 1, 3, and 4. However, we acknowledge that we relied on data multiplication repeatedly to achieve these results. Additionally, during the testing phase, we encountered some sentences that were present in the training phase but with a different dialect. Nonetheless, we found that this was not a significant obstacle for the system to identify, especially in cases where Arabic dialects are closely related to one another.

On the other hand, in some of our experiments, such as experiments 2, 5, 6, and 7, our system produced very poor results. Upon further analysis, we identified that the small volume of data in the training phase of the model was the main reason behind this outcome. While we acknowledge that data size is a crucial factor in the effectiveness of NMT systems, acquiring sufficient high-quality training data for NMT can be an arduous and costly process.

The success of an NMT system relies heavily on the quantity of available training data. Generally, greater amounts of training data result in improved translation quality. Nevertheless, gathering ample high-quality training data for NMT can be arduous, time-consuming, and costly.

To tackle the data size issue, researchers have explored various methods, including transfer learning. This technique involves pre-training an NMT model on a vast, general domain corpus and fine-tuning it on a smaller, domain-specific corpus to enhance its performance in that domain.

4.5 Conclusion

This chapter provides a comprehensive overview of NMT with a specific focus on our significant contributions in the field of Arabic dialect translation. We delve into the available datasets and discuss the extensive research conducted in the Arabic dialect NMT domain.

In addition, we present our approaches to Arabic dialect translation by employing a range of NMT models. We delve into the intricacies of seq2seq, attentional seq2seq NMT, and Transformer models, which have gained substantial popularity in MT and have been extensively studied to evaluate their effectiveness in improving translation quality. These models have proven to be instrumental in advancing the field, showcasing their potential to elevate translation accuracy and fluency.

Nevertheless, throughout our research journey, we encounter a formidable challenge posed by the limited availability of data in Arabic dialects, particularly in the case of the Algerian dialect. The scarcity of suitable training data significantly impacts the performance of NMT models, and even with the adoption of advanced architectures such as the powerful Transformer model, we face difficulties in achieving the desired level of translation quality.

5

Transfer Learning Neural Machine Translation Overview and Contribution

5.1 Introduction

This chapter provides an overview of transfer learning in NMT and its potential for improving the accuracy of MT systems. Transfer learning is a widely adopted technique that enables the use of pre-trained models to perform tasks that are related or different. This technique has shown significant promise in NMT, particularly in addressing the challenges of translating low-resource languages, such as Arabic dialects.

The chapter is divided into four sections: transfer learning, transfer learning NMT overview, transfer learning NMT contribution, and conclusion. In section 5.2, we discuss transfer learning and its setting, including its different approaches and how it works. section 5.3 focuses on the work applied to transfer learning in the context of improving NMT performance for Arabic dialects.

Section 5.4 is the most significant section of the chapter, where we present our contribution to the field of Arabic dialect NMT. Specifically, we discuss how we applied transfer learning to improve the accuracy of Arabic NMT and present our proposed approach in this context. And, section 5.5 provided a conclusion to this chapter.

In summary, this chapter highlights the potential of transfer learning in NMT and its ability to improve machine translation systems' accuracy. Our contribution in this chapter provides a potential solution to the significant challenge highlighted in the previous chapter,

which is the limited availability of data for Arabic dialects, particularly the Arabic dialect. By applying transfer learning to NMT, we were able to improve the accuracy of Arabic NMT while using a smaller training set.

5.2 Transfer Learning

In less than ten years, NMT has experienced enormous progress, has entered a mature phase [84], and has become widely used as compared to earlier traditional translation approaches. Despite its wide polarization, it produces generally poor results with resource-limited language pairs. Furthermore, creating high-resource parallel corpora from scratch is difficult, expensive, and time-consuming. As a consequence, the implementation of NMT approaches for low-resource language pairs has received a lot of attention in recent NMT research, resulting in a significant amount of research on this topic.

Some research exploited data from auxiliary languages to improve MT performance for resource-limited language pairs, and this last one contains three strategies, Multilingual Training, Transfer Learning, and Pivot Translation strategies. Multilingual NMT is a translation framework that uses more than one language pair [85, 86], training multiple language pairs in a single model via parameter sharing [87–91] due to the model’s ability to obtain a shared semantic representation between languages [92]. Moreover, the Transfer learning framework involves training a model on high-resource language pairs or multi-language pairs and transferring the knowledge learned previously to translate new low-resource language pairs [93–95]. However, the pivot translation framework is dependent on the availability of a high-resource intermediate language (Pivot) [96–99]. The translation of the source language to the target language is deconstructed in pivoting into the task of training two high-resource independent models: source-to-pivot languages and pivot-to-target languages. To obtain the target sentence, a source sentence is first translated using the source-pivot translation model, and the output is then translated using the pivot-target translation model.

Within this thesis, we make use of the transfer learning framework. The description and categorization of learning transfer are based on the explanation provided in the paper by Pan, S. J., and Yang, Q [100], notation and formulas utilized are also taken from the same paper.

Transfer learning involves extracting knowledge from source tasks and applying it to a target task. Unlike the conventional approach of simultaneously learning all source and target tasks, transfer learning focuses specifically on enhancing performance in the target task. This technique aims to transfer acquired knowledge from one or multiple source tasks to a designated target task. In contrast to traditional learning methods that tackle both source and target tasks simultaneously, transfer learning introduces a different emphasis.

In the context of translation, transfer learning relies on leveraging resources effectively.

It entails transferring knowledge from a source language with ample linguistic resources to a target language with limited resources. For this knowledge transfer to be effective, it requires either linguistic overlap or a substantial linguistic connection between the two languages. Overlap occurs when a language with limited resources is embedded within a linguistically richer language, as seen with the Algerian dialect within the broader Arabic dialects. A linguistic connection is evident when languages share commonalities such as shared letters or similar grammatical structures, as seen in the case of English and French.

Typically, a language translation is consisting of a domain D and a task T , where:

- Domain $D = \{\bar{X}, P(X)\}$ where the first component is the feature space $\bar{X} = \{x_1, x_2, \dots, x_n\}$ and the second component is the marginal distribution of probability $P(X)$.
- Task $T = \{\bar{Y}, f(\cdot)\}$ where the first component is the label space $\bar{Y}$ and the second component is an objective predictive function $f(\cdot) \approx P(y|x)$ with $x \in \bar{X}$ and $y \in \bar{Y}$.

In SL translation, x_i is the i th word vector corresponding to a sentence in the SL, then $\bar{X}$ is the space of all word vectors, and X is the particular learning sample (source sentence), furthermore, where $\bar{Y}$ is the label space, y_i is the i th label word vector corresponding to sentence in the TL.

The objective of transfer learning is to improve the learning of the target predictive function $f_{target}(\cdot)$ using the knowledge in SL , when $D_{source} \neq D_{target}$, or $T_{source} \neq T_{target}$. However, the scenario where $D_{source} \neq D_{target}$ implies that $\bar{X}_{source} \neq \bar{X}_{target}$ or $P_{source}(x) \neq P_{target}(x)$. Similarly, when $T_{source} \neq T_{target}$ implies $\bar{Y}_{source} \neq \bar{Y}_{target}$ or $P(y_{source}|x_{source}) \neq P(y_{target}|x_{target})$. Theory, when $D_{source} = D_{target}$ and $T_{source} = T_{target}$ the learning problem was traditional MT problem, otherwise, when $D_{source} \neq D_{target}$ or $T_{source} \neq T_{target}$ there are four learning problem cases:

- When the features spaces between the domains are different $\bar{X}_{source} \neq \bar{X}_{target}$ and marginal probabilities are the same $P_{source}(x) = P_{target}(x)$, corresponds that the languages to be translated were different from each other.
- When the features spaces between the domains are the same $\bar{X}_{source} = \bar{X}_{target}$ but the marginal probabilities are different $P_{source}(x) \neq P_{target}(x)$, may corresponds that the two languages to be translated focused in different topics.
- When the label spaces between the tasks are different $\bar{Y}_{source} \neq \bar{Y}_{target}$ and the objective functions are the same $P(y_{source}|x_{source}) = P(y_{target}|x_{target})$, corresponds that the languages to be translated into were different from each other.
- When the label spaces between the tasks are same $\bar{Y}_{source} = \bar{Y}_{target}$ but the objective functions are different $P(y_{source}|x_{source}) \neq P(y_{target}|x_{target})$, corresponds to the situation where the frequency of certain words in the target language is lower, resulting in challenges in accurately translating sentences that contain those specific words.

All the previous cases indicate that there must exist some relationship, explicit or implicit, between the two domains or tasks to be related to applying the transfer learning framework.

To conclude, the transfer learning aims to improve the learning of the target predictive function $f_{target}(\cdot)$ in D_{target} using the knowledge in D_{source} and T_{source} , where $D_{source} \neq D_{target}$ or $T_{source} \neq T_{target}$. To improve it, Pan, S. J., and Yang, Q [100] categorize the transfer learning to three main settings.

5.2.1 Transfer Learning Setting Categorization

Pan, S. J., and Yang, Q [100] set three essential questions, as outlined in [100], to assist in identifying the suitable transfer learning category for a given problem, 'What to transfer', 'How to transfer', and 'When to transfer'. The predominant focus in transfer learning studies is on addressing the aspects of 'What to transfer' and 'How to transfer'.

When different domains are related, there is common knowledge, this Knowledge shared between different domains helps to improve the performance of the target domain or task, thus transfer learning becomes effective. However, transfer learning can be unsuccessful, which is referred to as negative transfer when the source domain and the target domain are unrelated to each other.

Transfer learning searches to discover which knowledge can be transferred, especially, discovering which part of knowledge is across domains or tasks, which corresponds to the 'what to transfer' issue. Then, in order to transfer the knowledge, learning algorithms must be developed, which corresponds to the 'how to transfer' issue. Therefore, Pan, S. J., and Yang, Q [100] categorize transfer learning into three settings: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning. In inductive transfer learning, the source and target domains are the same, but the target task differs from the source task. In transductive transfer learning, the source and target domains differ, but the target task is the same as the source task. In contrast, in the unsupervised transfer learning scenario, both the target domain and task are different from, yet related to, the source domain and task.

5.2.1.1 Inductive Transfer Learning

Given a source domain D_{source} and a learning task T_{source} , as well as a target domain D_{target} and a learning task T_{target} , inductive transfer learning attempts to improve in the learning of the target predictive function $f_{target}(\cdot)$ in D_{target} using the knowledge in D_{source} and T_{source} , where $T_{source} \neq T_{target}$. The inductive transfer learning problem is categorized into two categories according to the availability of labeled data:

- When there is a large amount of labeled data in the source domain. Thus, the inductive transfer learning simply strives to achieve high performance in the target task by

transferring information from the source task.

- When there is no labeled data in the source domain. As a result, the label spaces between the source and target domains can differ, implying that the source domain side of information cannot be used directly.

5.2.1.2 Transductive Transfer Learning

The initial mention of transductive transfer learning can be attributed to Arnold et al [101]. its scenario involves a source domain D_{source} and a learning task (T_{source}), beside a target domain D_{target} and a corresponding learning task T_{target} . The objective of transductive transfer learning is to enhance the learning of the target predictive function $f_{target}(\cdot)$ in D_{target} using the knowledge in D_{source} and T_{source} , where $D_{source} \neq D_{target}$ and $T_{source} = T_{target}$. In addition, some unlabeled target domain data must be available at training time. Furthermore, as $D_{source} \neq D_{target}$, this setting can be split to two cases:

- The feature spaces between the source and target domains are different $\bar{X}_{source} \neq \bar{X}_{target}$.
- The feature spaces between domains are the same $\bar{X}_{source} = \bar{X}_{target}$, but the marginal probability distributions of the input data are different $P_{source}(x) \neq P_{target}(x)$.

5.2.1.3 Unsupervised Transfer Learning

In the context of unsupervised transfer learning, the target task differs from the source task but maintains a relationship, much like the inductive transfer learning scenario. Nevertheless, in the unsupervised transfer learning scenario, both the source and target domains lack labeled data during the training process.

Unsupervised transfer learning aims to help improve the learning of the target predictive function $f_{target}(\cdot)$ in D_{target} using the knowledge in D_{source} and T_{source} , where $T_{source} \neq T_{target}$ and $\bar{Y}_{source}$ and $\bar{Y}_{target}$ are not available, given a source domain D_{source} with a learning task T_{source} , a target domain D_{target} and a corresponding learning task T_{target} .

5.2.2 Approaches to Transfer Learning Setting

Based on 'How to transfer', there are four categories of transfer learning algorithms:

- Instance-based algorithms transfer knowledge by attaching weights to source instances.
- Feature-based algorithms transfer knowledge by spanning the subspace of the features in both source and target domains.

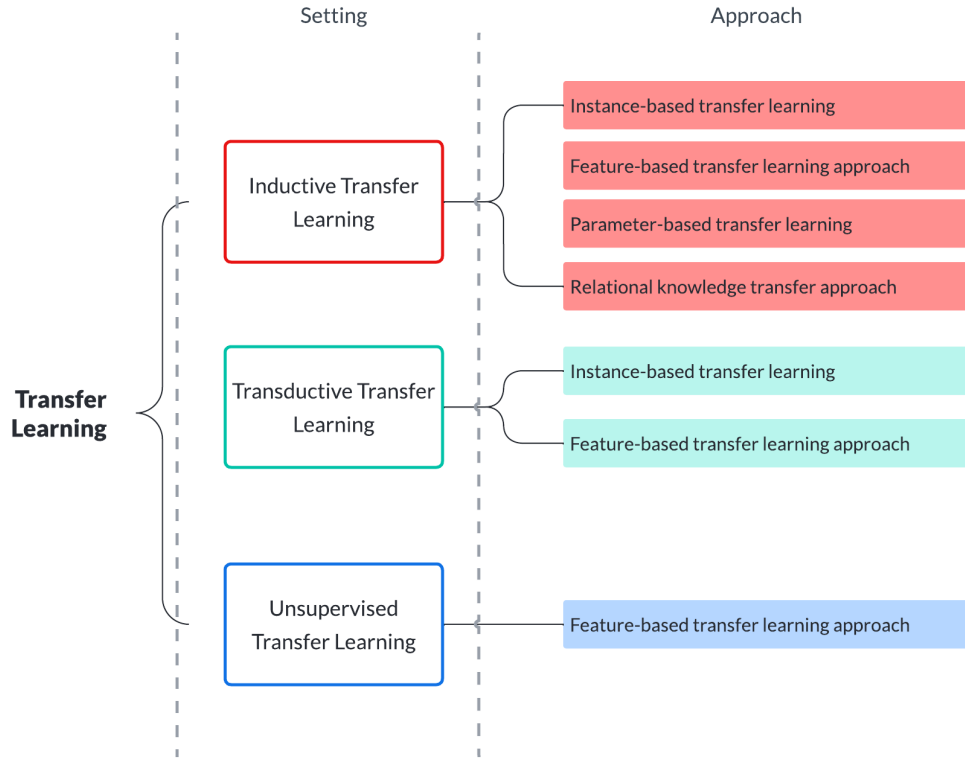


Figure 5.1: The suitable approaches for transfer learning settings.

- Parameter-based algorithms transfer knowledge embedded in part of the source domain models.
- Relation-based algorithms transfer knowledge by specifying rules between entities in the source domain.

Instance-based, feature-based, and parameter-based transfer learning methods are grouped together, all of which make the implicit assumption that data instances are independently and identically distributed. On the other hand, in the relational-based approach targets domains the data instances have inherent structures [102]. Additionally, each transfer learning approach is suited to different settings, and may not be effective for others, As illustrated in Figure 5.1.

5.2.2.1 Instance Based Transfer Learning

Instance-based transfer learning aims to resolve differences in marginal distribution by adjusting the weights of samples in the source domain [103–105]. These weighted samples are then utilized for training in the target domain, allowing the target learner to utilize only the relevant information from the source domain.

Instance transfer approach to inductive transfer learning setting, is applied using algorithms and methods such as TrAdaBoost algorithm, Heuristic method, and active learning.

TrAdaBoost [106] bases its assumptions on the notion that the source and target domains data use the exact same set of features $\bar{X}_{source} = \bar{X}_{target}$ and labels $\bar{Y}_{source} = \bar{Y}_{target}$, but the distributions of the data in the two domains are different. Furthermore, certain data from the source domain may be helpful in learning for the target domain, while others may not due to the different distributions between the source and target domains. For that, TrAdaBoost tries to iteratively reweight the data from the source domain to lessen the impact of the negative source data and to boost the positive source data contribution to the target domain.

While, heuristic method [107] based on the distinction between conditional probabilities $P(y_{source}|x_{source})$ and $P(y_{target}|x_{target})$ to exclude harmful training samples from the source domain. Active learning method [107] select the unlabeled data in a target domain to be labeled using source domain data.

$$\theta* = \underset{(x,y) \in D_{target}}{\operatorname{Argmin}} \sum P(D_{target})l(x,y, \theta) \quad (5.1)$$

On other hand, the instance transfer approaches to transductive transfer learning problem, aims to learn an optimal model for the target domain by minimizing the expected risk (Formula 5.1, where $l(x,y, \theta)$ is a loss function that depends on the parameter θ , and since no labeled data in target domain, the model needs to be learned from the source domain. In this scenario, two cases emerge:

- When $P(D_{source}) = P(D_{target})$, the model can be learn from the source domain data instead target domain data, using the formula 5.2.

$$\theta* = \underset{(x,y) \in D_{source}}{\operatorname{Argmin}} \sum P(D_{source})l(x,y, \theta) \quad (5.2)$$

- When $P(D_{source}) \neq P(D_{target})$, the model must be learned with a high generalization ability for the target domain. Consequently, the optimization problem's formulation is modified to that presented in formula 5.3. The transductive transfer learning problem can be solve by estimate $\frac{P(x_{source_i})}{P(x_{target_i})}$ for each instance.

$$\theta* = \underset{(x,y) \in D_{source}}{\operatorname{Argmin}} \sum \frac{P(x_{source_i})}{P(x_{target_i})} l(x_{target_i}, y_{source_i}, \theta) \quad (5.3)$$

In order to estimate $\frac{P(x_{source_i})}{P(x_{target_i})}$ directly, Sugiyama et al [108] proposed the Kullback-Leibler Importance Estimation Procedure (KLIEP) algorithm, which is based on the minimization of the Kullback-Leibler divergence, which can be integrated with cross-validation to estimat the weights of the source domain data, and training the model on the reweighted data. Besides reweighting techniques, there are several methods used in regression and classification to estimate $P(x_{source_i})$ and $P(x_{target_i})$.

5.2.2.2 Feature Based Transfer Learning

In practice, certain words or features may be present in one domain but absent in another. Consider a scenario where one domain represents the Algerian dialect and another the Egyptian dialect, both being subsets of Arabic dialects. While these share a common feature space, specific features might be unique to each. For instance, the term 'براف' (meaning 'a lot' in Algerian dialect) exists in the Algerian dialect but not in the Egyptian one. This demonstrates domain-specific features.

Feature-based transfer learning techniques enable knowledge transfer in specific feature spaces rather than raw input spaces. These methods require matching labels ($y_{source} = y_{target}$) and some overlapping data ($\bar{X}_{source} \cap \bar{X}_{target} \neq \emptyset$).

Fundamentally, feature-based transfer learning methods operate on the premise of acquiring a set of mapping functions. These functions serve to convert data from both the source and target domains into a shared feature space, thereby minimizing the differences between the domains. Subsequently, a target model is trained within this new feature space, utilizing data sourced from the mapped versions of the source and target domains [102].

For the inductive transfer learning scenario, the feature-based transfer learning approach can use either supervised learning or unsupervised learning, depending on the availability of labeled data :

- The objective of supervised learning in the inductive transfer learning is to acquire a low-dimensional representation that can be shared among related tasks. This is achieved by solving an optimization problem, represented as equation 5.4, where the tasks in the source and target domains are represented by S and T , respectively. The optimization problem estimates the model parameters A , and the low-dimensional representations $U^T X_T$ and $U^T X_S$, simultaneously. The matrix $A = [a_S, a_T] \in \mathbb{R}^{d \times 2}$ is a set of parameters, while U is a orthogonal matrix $d \times d$, which represents the mapping function for transforming the original high-dimensional data into low-dimensional representations. The (r, p) - norm of A is given by $\|A\|_{r,p} := \left(\sum_{i=1}^d \|a^i\|_p^r \right)^{\frac{1}{p}}$.

$$\begin{aligned} \underset{U, A}{\operatorname{argmin}} \sum_{t \in (T, S)} \sum_{i=1}^{n_t} L(y_t, \langle a_t, U^T x_{t_i} \rangle) + \gamma \|A\|_{2,1}^2 \\ \text{s.t. } U \in \mathbf{O}^d \end{aligned} \quad (5.4)$$

- In order to enable inductive transfer learning in unsupervised methods, various strategies have been put forth. Among these strategies is one proposed by [109], which leverages sparse coding, which serves as a technique for constructing features that are capable of learning more advanced features in a hierarchical manner.

For the transductive transfer learning scenario, several methods have been proposed. [110] proposed a kernel-mapping function that maps data from both the source and target domains to a high-dimensional feature space. The models are then trained using standard discriminative learning methods. [111] suggested transfer learning via dimensionality reduction. They used the Maximum Mean Discrepancy Embedding (MMDE) approach to reduce the dimensionality of the data and learn a low-dimensional space that reduces the difference of distributions between different domains.

For the unsupervised transfer learning scenario, where there is an absence of labeled data in both the source and target domains, it's evident that there is limited research in this area. Notably, [112] proposed a self-taught clustering algorithm that uses a large set of unlabeled data in the source domain to cluster a small set of unlabeled data in the target domain. This can be used to provide pseudo-labels for the target data, which can then be used to train a classifier. And, [113] proposed a transferred discriminative analysis algorithm that uses dimensionality reduction techniques to reduce the dimensionality of the target data and labeled source data. The reduced data is then used to train a classifier.

5.2.2.3 Parameter Based Transfer Learning

In the context of inductive transfer learning, the paradigm of parameter-based transfer learning, also known as model-based transfer learning, operates under the assumption that there exists shared knowledge between the source task and the target task at the model level. This implies that the knowledge transferred is embedded within the model parameters, priors, or even the model architectures themselves. It's important to note that this method requires the presence of certain labeled instances within the target domain. [102] categorized various model-based transfer learning approaches into two distinct groups:

1. The transfer of knowledge through shared model components and regularization encompasses transfer learning techniques that involve constructing the target model by reutilizing specific components or hyperparameters from the source model [114–116].
2. Transferring knowledge through regularization involves the application of constraints on parameters based on specific prior hypotheses. Regularization is a method for solving poorly posed machine learning issues and avoiding model overfitting by limiting model flexibility [102].

The regularization framework is designed to work under multitask learning. Multitask learning attempts to learn both the source and target tasks concurrently, as was described in Section 5.2, whereas transfer learning aims to improve the performance of the target domain using the data from the source domain. In a more intuitive sense, considering a well-trained source model θ_{source} that has absorbed a substantial amount of structural

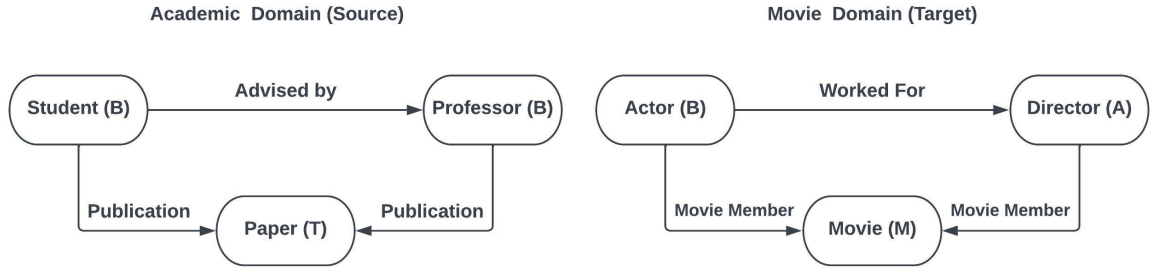


Figure 5.2: Relational Information Transfer from the Academic to Cinematic Domain. This figure illustrates the process of transferring relational information from the academic domain to the cinematic domain. The left side shows a knowledge graph representing academic concepts, while the right side shows a knowledge graph representing cinematic concepts. The arrows between the two graphs indicate the transfer of relational information using a transfer learning approach. The approach allows the model to leverage the knowledge learned in one domain to improve performance in another domain [102, 117].

knowledge from the data, this acquired structure can be effectively transferred to refine the learning process for a more accurate target model θ_{target} , especially when dealing with low-resource target tasks. Hence, the approach of reusing the model trained on the source domain proves to be a more effective strategy.

5.2.2.4 Relational Knowledge Based Transfer Learning

The approach of relational knowledge transfer into the inductive transfer learning scenario pertains to the application of transfer learning in domains that are characterized by relations, encompassing scenarios where data can be expressed through multiple interconnections. This approach operates under the fundamental presumption that there exists a relational similarity between the data instances within both the source and target domains. Consequently, the aim of this approach is to facilitate the transfer of these relational connections from a source domain to a target domain. Figure 5.2 provides an illustration of how to transfer relational information from the academic domain to the cinematic domain, which was taken from [102, 117].

As:

$$P_1(x, y) \wedge P_2(x, z) \implies P_2(y, z) \quad (5.5)$$

Where P_1 and P_2 are relations between objects x, y and z . For academic domain, the relationship between the student and the professor can be represented as *AdvisedBy(student, professor)* and between student and his paper can be presented as *Publication(student, paper)*, and Through the formula 5.5, the relationship between professor and paper was:

$$\begin{aligned} & \text{AdvisedBy}(\text{Student}, \text{Professor}) \wedge \text{Publication}(\text{Student}, \text{Paper}) \\ & \implies \text{Publication}(\text{Professor}, \text{Paper}) \end{aligned} \quad (5.6)$$

And the same relations for cinematic domain:

$$\begin{aligned} & \text{WorkedFor}(\text{Director}, \text{Actor}) \wedge \text{MovieMember}(\text{Actor}, \text{Movie}) \\ & \implies \text{MovieMember}(\text{Director}, \text{Movie}) \end{aligned} \quad (5.7)$$

In this context, [118] suggested Markov logic networks, a logic probability mixed model in which relations are encoded as predicates and the regularities of relations are represented as formulas. Relationship-based transfer learning approaches using Markov logic networks first learn accurate weighted logic formulations that express the relationship regularities from the source domain. Then, they generate candidate logic formulas with predicates from the target domain based on these regularities. Finally, they sift, revise, and reweight these candidates to accurately represent the target domain.

5.3 Arabic Transfer learning NMT: an Overview

In the previous chapter, we tackled the problem of NMT for Arabic dialects, impeded by limited data availability. Consequently, Researchers have explored strategies like inductive transfer learning [119–121], and unsupervised transfer learning [122] to enhance NMT model performance for Arabic dialects. A comparative analysis of these approaches is presented in Table 5.1.

[119] proposes an intermediary decoder that connects two encoder-decoder models. The first encoder-decoder model translates MSA into English, while the second encoder-decoder translates Arabic dialects into MSA. The proposed model leverages information from various connected tasks, and it may learn to construct sentences in the correct target language order, resulting in more fluent and accurate translations.

[120] employs an NMT system for six Arabic dialects using multitasking learning to address the lack of parallel data for these vernacular languages. The models are trained on a small corpus (PADIC) containing only 6,400 parallel sentences for each dialect. The multitasking approach benefits from the similarities between Arabic dialects, enabling it to outperform single-task learning and SMT method.

[121] proposes a unified multitask NMT model's training process, which involves simultaneous training on various tasks, including translation and part-of-speech tagging. This model uses a shared representation, leveraging inductive knowledge transfer between

Table 5.1: Related works of Transfer Learning strategy applied on Arabic dialect NMT.

Work	Year	Learning Strategy	Model	Dataset	Train data size	Dialect	Bleu Score
[119]	2018	Inductive	Two seq2seq model connected with shared decoder	PADIC And MPCA	17736	Maghreb - MSA	41.00
[120]	2022	Inductive	Attentional seq2seq with multiple decoder	PADIC	13805	Levantine - MSA	30.00
					6400	ALG - MSA	35.06
					6400	ANB -MSA	30.81
					6400	Tunisian -MSA	24.42
					6400	Palestine - MSA	34.50
[121]	2018	Inductive	Seq2seq model with multiple shred layers	PADIC And MPCA	6400	Syria - MSA	27.85
					6400	Moroccan - MSA	26.18
					13805	Levantine - MSA	41.00
					17736	Maghrebi - MSA	30.00
					143,518	Saudi - MSA	31.06
[122]	2020	Unsupervised	Attentional Sequence-to-Sequence Translator with residual connections and a bi-directional layer	Dataset designed by the authors	149,582	Egyptian - MSA	24.42

tasks to enhance performance. The architecture employs an RNN encoder-decoder architecture, with separate encoders and decoders for distinct translation tasks (Arabic dialect-MSA and MSA-English) while sharing private and invariant layers. The model can leverage the characteristic information and improve translation quality from Arabic dialects to MSA by alternately training the translation tasks and part-of-speech tagging tasks.

[122] develops an unsupervised approach to Arabic language learning using popular words in both MSA and Arabic dialects. The goal is to find a translation matrix that maps from one hyperdimensional representation for dialectal and MSA through common anchor words and tackle out-of-vocabulary issues with subword units. The unsupervised technique is tested on two models, GNMT and attentional seq2seq.

Our approach to improving NMT for Arabic dialects involved the use of transductive transfer learning. Although previous studies have investigated different transfer learning strategies to address the issue of limited parallel corpora for Arabic dialects, we have not come across any prior work that has employed transductive transfer learning for this purpose. Our study aims to address this gap and investigate the potential of this technique for enhancing machine translation performance for Arabic dialects.

5.4 Algerian Dialect Transfer Learning NMT: a Contribution

Unfortunately, Arabic dialects, including the Algerian dialect, suffer from a severe shortage of high-quality parallel texts [123, 124]. The scarcity of parallel transcripts poses a major challenge in developing reliable and efficient NMT models, as elucidated by the findings presented in the preceding chapter. Collecting relevant and high-quality parallel corpora is a daunting task that requires a considerable amount of time and resources. Moreover, the quality of parallel corpora can vary widely, depending on factors such as domain, dialect, and textual style, further complicating the collection process.

To address this issue, we have turned to transfer learning as a viable strategy for improving the results of low-resource Algerian dialect. By transferring knowledge from a sizable source corpus which is Arabic dialects to a smaller target corpus which is the Algerian dialect, transfer learning enables the model to leverage the knowledge and structure learned from all Arabic dialects, thereby improving the results on the Algerian dialect. We made the decision to transfer knowledge from Arabic dialects to Algerian dialect based on several underlying reasons, which include:

- Numerous Arabic dialects, including Algerian, possess shared linguistic characteristics such as in vocabulary, syntax, and grammar. Transfer learning leverages these commonalities to enable the model to transfer knowledge from other Arabic dialects to Algerian dialect.

- When the model begins the learning process with the shared features found across all Arabic dialects and then specializes in learning the specific characteristics of the Algerian dialect, it leads to rapid convergence and improved overall performance.

Furthermore, As previously noted in Chapter 1, Arabic dialects are classified into distinct groups, forming a complex mosaic of linguistic diversity. Within each group of Arabic dialects, there are noticeable linguistic resemblances that differentiate them from dialects belonging to other groups. To illustrate, dialects within the Maghrebi group, spanning North African nations, showcase a spectrum of shared linguistic traits, encompassing the incorporation of Berber loanwords, distinctive phonological features, and common syntactic structures. Conversely, when we shift our focus to dialects from different groups, the presence of linguistic disparities becomes evident. For instance, Gulf dialects stand apart considerably from Maghrebi or Levantine dialects, boasting their own lexicon, pronunciation nuances, and grammatical structures.

Nevertheless, while we acknowledge the presence of shared linguistic features among these dialects, it's essential to emphasize that the degree of similarity varies. Some dialects exhibit a higher percentage of similarity, while others show a lower degree of similarity. This led us to contemplate the possibility of achieving improved outcomes through an incrementally transfer the knowledge. In which, we can strategically implement transfer learning by initially transferring knowledge from the larger and more diverse group of Arabic dialects to the Maghreb dialects, and subsequently, to the Algerian dialect. This sequential transfer of knowledge allows for a gradual learning of the NMT model. Starting with the vast group of Arabic dialects provides a robust foundation, which is then will specialize on the specific linguistic nuances of the Maghreb dialects. From there, this refined knowledge is further will specialize on the distinctive characteristics of the Algerian dialect. This incremental transfer learning approach offers a systematic way to enhance the model's performance, leveraging the linguistic features and differences across these dialect groups to achieve superior results in translating the Algerian dialect.

Within this section, we present two approaches aimed at enhancing the performance of the NMT model in translating the Algerian dialect. The initial approach involves employing the transfer learning strategy (section 5.4.1), while the second entails implementing our proposed incremental learning transfer strategy (section 5.4.2). In order to validate the effectiveness of these strategies, we extended their application to address the challenges posed by low-resource Tunisian and Moroccan dialects.

5.4.1 Transductive Transfer Learning

In our research, we focused on enhancing NMT model performance for translating the Algerian dialect. This improvement was accomplished by applying transfer learning, leveraging the knowledge from different Arabic dialects, as detailed in [80]. However, during

this effort, we confronted two critical questions that required thorough investigation:

Firstly, we determine which specific knowledge to transfer to enhance NMT performance. We do this by exploring linguistic similarities between Arabic dialects and the Algerian dialect (Section 5.4.1.1).

Secondly, we find the most efficient strategy to transfer this knowledge. This involves selecting the optimal transfer learning framework and approach tailored to our unique problem (Section 5.4.1.2).

5.4.1.1 Linguistic Background

While the Algerian dialect is part of the broader Arabic dialects, it shares numerous similarities, which can be distilled into several key points.

In terms of lexicology, when comparing Arabic dialects with the Algerian dialect, there are several noteworthy aspects to consider. Firstly, both Arabic and Algerian dialects trace their linguistic roots back to a common ancestor, Classical Arabic. This classical language served as a literary standard across the Arab world, encompassing various lexical categories such as verbs, nouns, pronouns, and particles. This shared linguistic heritage has resulted in a significant overlap in vocabulary between different Arabic dialects, including Algerian Arabic. Secondly, due to historical and cultural connections between Algeria and other Arab nations, there has been substantial linguistic exchange between the Algerian dialect and other Arabic dialects. Consequently, numerous loanwords and borrowings have enriched the vocabularies of these dialects.

To illustrate this point, we created a heatmap, as presented in the Figure 5.3, to display the similarities and dissimilarities between the Algerian dialect and other Arabic dialects. The purpose of this was to showcase the potential transferability of certain lexical features from other Arabic dialects to Algerian dialects. A heatmap was generated using the MADAR and PADIC datasets to highlight the shared vocabulary between the Algerian dialect and other Arabic dialects. The Algerian dialect data was sourced from the PADIC dataset, while the MADAR dataset provided data on other Arabic dialects. The heatmap clearly indicates that the Algerian dialect shares numerous words with other Arabic dialects.

When it comes to phonology, the Algerian dialect stands out with distinct sound patterns in comparison to MSA. However, it shares some commonalities with other Arabic dialects. One of the most significant differences lies in the pronunciation of the consonant ‘ق’. While MSA typically uses /q/, the Algerian dialect showcases a range of variations, including /q/, /g/, /k/, and even /ʔ/. The /g/ sound is reminiscent of Gulf and Iraqi dialects, while /k/ and /g/ are characteristic of the Moroccan and Tunisian dialects. The /ʔ/ pronunciation, often denoting a glottal stop, is a distinctive feature of the Egyptian dialect [125–127].

	بابسو	حطيتها	المكتوب	مخدوم	الدقيق	الفطور	عايش	الناس	بلاصتي	سببطار
Aleppo	1	0	1	0	0	1	0	1	0	0
Alexandria	1	0	1	0	0	0	1	1	0	0
Algiers	1	1	1	1	1	1	1	1	1	1
Amman	1	0	1	0	0	1	1	1	0	0
Aswan	1	0	1	0	0	0	1	1	0	0
Baghdad	0	0	1	0	0	0	0	1	0	0
Basra	0	0	1	0	0	0	1	1	0	0
Beirut	1	0	1	0	0	0	1	1	0	0
Benghazi	1	0	1	0	0	1	0	1	0	0
Cairo	1	1	1	0	1	0	1	1	0	0
Damascus	0	0	1	0	0	1	1	1	0	0
Doha	1	0	1	0	0	0	0	1	0	0
Fes	1	0	1	0	0	1	0	1	0	0
Jeddah	1	1	0	0	0	1	0	1	0	0
Jerusalem	1	0	1	0	0	1	0	1	0	0
Khartoum	1	0	1	0	0	1	0	1	0	0
Mosul	1	0	0	0	1	0	0	1	0	0
Muscat	0	0	1	0	1	0	0	1	0	0
Rabat	1	0	1	0	1	1	0	1	1	1
Riyadh	0	0	1	0	0	1	0	1	0	0
Salt	0	1	1	0	0	1	0	1	0	0
Sanaa	0	0	1	0	0	0	1	1	0	0
Sfax	1	0	1	1	0	1	0	1	1	1
Tripoli	1	0	0	0	0	1	0	1	0	0
Tunis	1	0	1	0	0	1	0	1	1	1

Figure 5.3: Heatmap of Similarity of some words between Algerian and Other Arabic Dialects. The heatmap depicts shared vocabulary between Algerian dialect and other Arabic dialects, using green to indicate similarity and yellow to indicate dissimilarity. Constructed using MADAR and PADIC datasets, the results demonstrate that the Algerian dialect shares many words with other Arabic dialects, especially those spoken in North Africa and the Middle East.

Another notable phonological variation is the pronunciation of the MSA consonant ‘ث’ /θ/ in the Algerian dialect. In some cases, it may be pronounced as ‘ت’ /t/, resembling the Levantine and Egyptian dialects. For instance, the word ‘ثوم’ /θuwm/ in MSA becomes ‘توم’ /tuwm/ in the Algerian dialect [127].

Additionally, the pronunciation of the MSA consonant ‘ذ’ as /d/ is another significant distinction in Algerian, Levantine, and Egyptian dialects. In Tunisian and Egyptian dialects, similar to the Algerian dialect, the glottal stop ‘ء’ /ʔ/ can be elongated or omitted when it appears in the middle of certain words [127]. For instance, in the MSA word ‘ذئب’, the glottal stop may be elongated, resulting in its pronunciation as ‘ذيب’.

In terms of morphology, the Algerian dialect exhibits morphological features that diverge from MSA but align more closely with certain Arabic dialects [126]. These features include:

- The singular word declension, which corresponds to the nominative, genitive, and accusative cases at the end of the word, is not accepted in all Arabic dialects [128]. Also, the three doubled case endings that express nominal indefiniteness are also dropped.
- The nominal dual, a unique feature of MSA, is notably absent from the Algerian dialect, as it is in several other Arabic dialects [128].

- Several Arabic dialects, similar to the Algerian dialect, formed non-MSA negations by adding the prefix ‘ما’ and the suffix ‘ش’ to the verb [128–130].

At the syntactic level, the Algerian dialect shares several similarities with other Arabic dialects. One of the most notable syntactic similarities is the use of SVO word order in declarative sentences, which is common in many Arabic dialects, including Algerian. This word order differs from the SOV word order in MSA, which places the object before the verb.

At the lexical level, the common vocabulary and borrowed terms found in the Algerian dialect, along with its counterparts like Moroccan, Tunisian, and Egyptian Arabic, hold the potential to enhance the precision and efficiency of MT systems. By identifying and harnessing these shared words and phrases, NMT systems can generate translations that are more natural and fluent.

Likewise, the phonological and morphological similarities shared among the Algerian dialect and other Arabic dialects, including the pronunciation of specific consonants or the absence of nominal dual forms, can aid NMT systems in better capturing the nuances and intricacies of the Algerian dialect’s sound and structure. Additionally, the analogous syntactic structures and word order patterns in the Algerian dialect and other Arabic dialects can assist NMT systems in comprehending sentence meanings and contexts more accurately, thereby yielding more precise translations.

5.4.1.2 Method

We aim to leverage the knowledge acquired from Arabic dialects to optimize our NMT model’s performance in translating the Algerian dialect, employing a transfer learning strategy. In this context, Arabic dialects serve as our D_{source} , Algerian dialect as our D_{target} , and MSA as our T_{source} and T_{target} . In our problem, it’s evident that the source domain (D_{source}) differs from the target domain (D_{target}). However, the source and target tasks ($T_{source} = T_{target}$) remain consistent, representing a transductive transfer learning scenario (presented in section 5.2.1.2).

As elaborated in Section 5.4.1.1, there are indeed significant shared characteristics between the Algerian dialect and multiple Arabic dialects. However, it is crucial to acknowledge that there are also numerous areas of distinction among these dialects, each characterized by its own unique features. Figure 5.4 provides a clear illustration of the dissimilarity at the lexicon level between Arabic dialects from the MADAR corpus and the Algerian dialect from the PADIC corpus. This reinforces the notion that there are significant differences between these dialects, which must be taken into account when developing our NMT model.

Furthermore, these disparities imply that the feature spaces may be identical, denoted as $\bar{X}_{source} = \bar{X}_{target}$, between the source domain (D_{source}) and the target domain (D_{target}). In

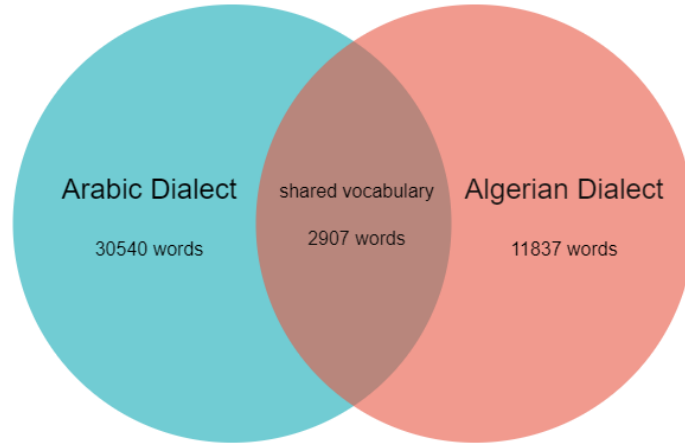


Figure 5.4: Lexical Dissimilarity between MADAR Arabic Dialects and Algerian Dialect from PADIC. This figure provides a visual representation of the dissimilarity at the lexical level between the Arabic dialects from the MADAR dataset (in blue) and the Algerian dialect from the PADIC dataset (in red), with the overlap of words between the two datasets in the middle. This figure emphasizes the importance of considering lexical differences when analyzing Arabic dialects.

such a scenario, the probability distributions of the input data for D_{source} and D_{target} may indeed differ, represented as $P_{source}(x) \neq P_{target}(x)$.

Understanding the relationship between source and target domains is pivotal when devising a knowledge transfer strategy within the realm of machine learning. In our specific case, we have established a substantial overlap in feature spaces between the source and target domains, as indicated by $\bar{X}_{source} \cap \bar{X}_{target} \neq \emptyset$. Furthermore, we have assumed that the labels in both domains are identical, $y_{source} = y_{target}$. Given these conditions, the instance-based transfer learning technique emerges as the most suitable approach for our scenario. This method aims to address disparities in marginal distribution by adjusting the weights of samples in the source domain. These weighted samples are subsequently leveraged for training in the target domain, enabling the target learner to extract pertinent information exclusively from the source domain.

Once we've established the overlap between the feature spaces of the source and target domains and assumed that the labels in both domains are equivalent, we employ the instance-based transfer learning approach to enhance the NMT model for the low-resource Algerian dialect into MSA, leveraging knowledge from various Arabic dialects. This transductive transfer learning strategy comprises two key steps, clearly depicted in 5.5.

In the initial step, the model is trained to estimate the weights of the source domain, comprehensively capturing and learning all the distinctive features of Arabic dialects. This process yields a model that specializes in translating all Arabic dialects, referred to as the parent model.

Subsequently, the second step entails training the pre-trained model (parent model) on the Algerian dialect. This training leads to the extraction of the child model, which

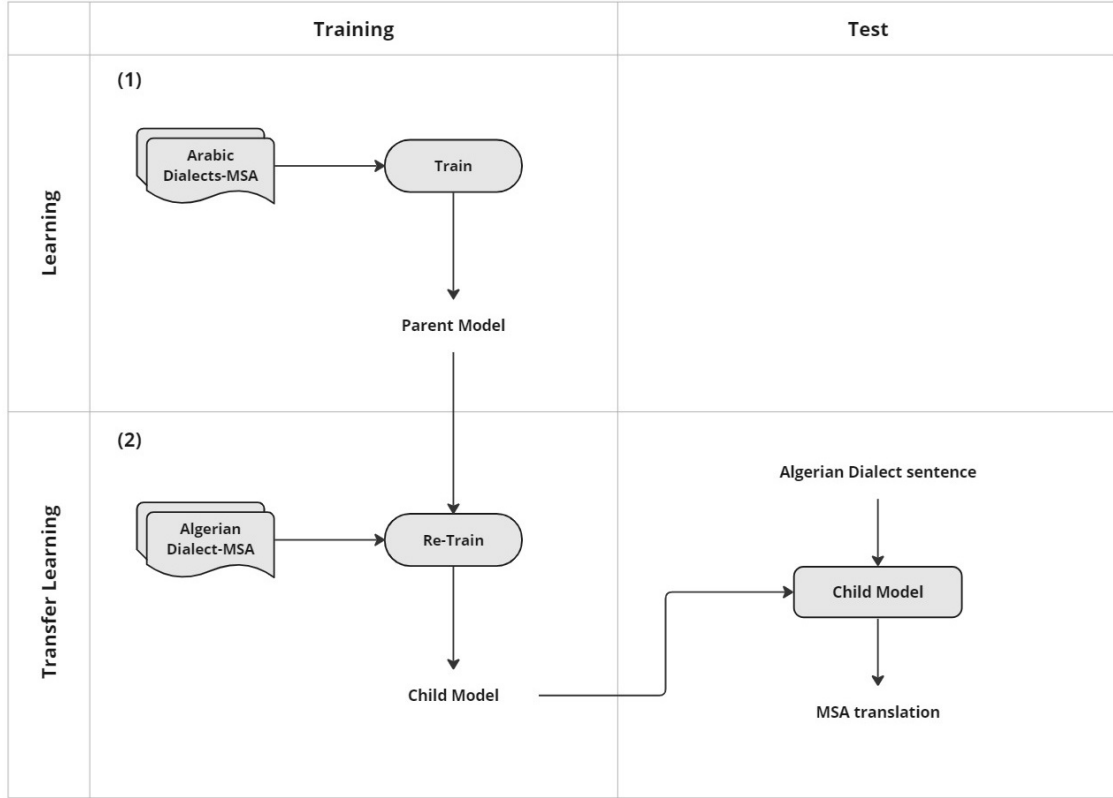


Figure 5.5: The Architecture of the transductive transfer learning strategy used.

specializes in translating Algerian dialects into MSA. Initially, the parent model is capable of translating the Algerian dialect, but it may occasionally introduce errors or substitute terms from other Arabic dialects due to the vocabulary containing terms from all Arabic dialects.

However, after fine-tuning the model with Algerian dialect data, it becomes highly skilled in understanding the specifics of the Algerian vocabulary. This specialization enables it to translate the Algerian dialect with much greater accuracy. By training the model on Algerian dialect data, we create a specialized child model optimized for translating the Algerian dialect into MSA.

We implemented our Instance-based transfer learning strategy to translate the low-resource Algerian dialect into MSA [80] using three models: Seq2Seq, Attentional-Seq2Seq, and Transformer models. All of these used models share the same architecture as detailed in the previous chapter.

5.4.1.3 Experimentation

In our experiments, we use transfer learning to improve the performance of NMT systems specifically for the low-resource Algerian dialect. We employed a range of model architectures, including seq2seq, attentional seq2seq, and transformer models, to carry out these experiments.

Our experimental design followed this sequence: initially, we trained the models using a diverse dataset encompassing all Arabic dialects. Subsequently, we fine-tuned these models by retraining them focusing on the low-resource Algerian dialect. Furthermore, to assess the efficacy of this strategy and validate its effectiveness, we extended its application to tackle the challenges posed by low-resource Tunisian and Moroccan dialects.

The main idea behind this strategy was to use a large dataset of Arabic dialects to build a model capable of accurately translating low-resource dialects. Since Algerian, Tunisian, and Moroccan dialects have similarities with various Arabic dialects, we initially trained our models on Arabic dialects to improve their performance in translating these low-resource dialects. To do this, we used the PADIC and MADAR corpora, which provide a substantial amount of data for various Arabic dialects. To optimize the utilization of our data, we conducted the following experiments:

Experiment 2-TL involved training the Seq2seq and attentional seq2seq models on 40000 sentences from the MADAR corpus, followed by retraining on 10000 Algerian dialect sentences from the PADIC corpus to obtain the child model. The performance of the child model was then assessed on 2729 Algerian dialect sentences from the same corpus [80].

Experiment 5-TL involved training the Attentional seq2seq and transformer models on 48000 sentences from the MADAR corpus, retraining on 11821 Algerian dialect sentences from the PADIC and MADAR corpora, and assessing the performance of the child model on 3000 Algerian dialect sentences from the same corpora.

In **Experiment 6-TL**, we trained the Attentional seq2seq and transformer models on 48000 sentences from the MADAR corpus, retrained it on 7411 Tunisian dialect sentences from the PADIC and MADAR corpora, and evaluated the child model on 3000 Algerian dialect sentences from the same corpora.

Lastly, **Experiment 7-TL** involved training the Attentional seq2seq and transformer models on 48000 sentences from the MADAR corpus, retraining it on 7411 Moroccan dialect sentences from the PADIC and MADAR corpora, and evaluating the child model on 3000 Algerian dialect sentences from the same corpora.

During Experiment 2-TL, the Algerian dialect data used contained two dialects, Annaba and Algiers, which could result in test data containing sentences from one dialect that appeared in the training data written in the other dialect. On the other hand, Experiments 5-TL, 6-TL, and 7-TL were designed so that the test data was not part of the training data, which ensured that the models were evaluated on their capability to generalize to target dialects.

We leveraged TensorFlow and Keras to implement our selected models. To evaluate the model's performance, we relied on accuracy metrics and categorical cross-entropy loss. The training process involved 50 epochs with each experiment, and we used a batch size of 512 sequences for each iteration. It's worth noting that we conducted several experiments

to arrive at the optimal batch size and the number of epochs, rather than choosing arbitrary values. Before beginning the training phase, we prepared the data by padding all sequences within a batch to match the length of the longest sequence. This approach ensured that the models could learn from all of the data without encountering issues such as vanishing gradients.

5.4.2 Incremental Transfer Learning: Proposed Strategy

In section 5.4.1, we focused on leveraging linguistic connections between the Algerian dialect and other Arabic dialects to enhance the Algerian dialect NMT system performance. We identified shared linguistic features that could be harnessed to improve NMT output accuracy and fluency.

Expanding upon the information presented, we have subsequently found more pronounced linguistic features that extend from the broader Arabic dialects family to the Maghreb dialects, and further specific features between the Maghreb dialects and the Algerian dialect. By transferring knowledge from Arabic dialects to Maghreb dialects and then from Maghreb dialects to the Algerian dialect, we can further enhance the performance of Algerian dialect NMT systems.

Our approach is data-driven and grounded in objective analysis and empirical evidence. Leveraging these linguistic features allows us to construct more precise and effective NMT models for the low-resource Algerian dialect.

In this section, we initially discuss the similarities between Arabic dialects and Maghreb dialects, followed by an examination of the shared features between Maghreb dialects and the Algerian dialect (Section 5.4.2.1). Then, we introduce our proposed incremental transfer learning strategy in Section 5.4.2.2.

5.4.2.1 Linguistic Background

This section will commence with a discussion of the common linguistic characteristics shared between Arabic dialects and Maghreb dialects. Subsequently, we will delve into the shared linguistic features between the Maghreb dialects and the Algerian dialect.

Linguistic Similarity Between Arabic Dialects and Maghreb Dialects:

At the lexical level, Maghreb dialects exhibit numerous resemblances to other Arabic dialects, notwithstanding variations in pronunciation and differences in vocabulary. For instance, words such as 'كتاب' (book), 'حجر' (stone), 'صباح' (morning), 'زيت' (oil), and 'دار' (house) are all employed in Maghreb dialects, just as they are in various other Arabic dialects. The lexical convergence between Maghreb dialects and the other Arabic dialects is

advantageous for the NMT model, enabling accurate translation. These shared linguistic features can enhance the model's performance and effectiveness.

At the phonological level, Maghreb dialects share many similarities with other Arabic dialects. For example, they use the same 28 Arabic consonant phonemes as standard Arabic. Additionally, they also use non-Arabic phonemes such as /g/, /p/, and /v/ which are borrowed from foreign languages.

At the morphological level, Maghrebi dialects exhibit similarities with other Arabic dialects. They've omitted certain unique features like the dual and feminine plural forms found in MSA. However, they share various morphological traits with Arabic dialects. For instance, similar to other Arabic dialects, Algerian, Moroccan, and Tunisian dialects don't utilize singular word declensions for the nominative, genitive, and accusative cases. Furthermore, the negation in Maghrebi dialects is formed by adding the prefix 'ما' and the suffix 'ش' to the verb, which is a feature shared by several Arabic dialects. This similarity in morphological structure highlights the close relationship between Maghrebi dialects and other Arabic dialects, as they share several common linguistic features that are not present in MSA.

Additionally, at the syntactic level, Maghrebi dialects share similarities in the word order of the sentence with other Arabic dialects. For instance, in Algerian, Tunisian, and Moroccan Arabic, the word order of a declarative sentence is quite flexible, but the SVO order is the most frequently used form since the two varieties have both SVO and VSO. This aligns with numerous other Arabic dialects where SVO also prevails as a prevalent word order.

In Section 5.4.1.1, we explored the lexical, phonological, morphological, and syntactic commonalities between Arabic dialects and the Algerian dialect. Although the Algerian dialect has unique features, it's important to note that it shares numerous similarities with other Arabic dialects. Notably, when compared to other Arabic dialects, the Algerian dialect exhibits a closer resemblance to the Tunisian and Moroccan dialects. This highlights the strong connection between the Maghrebi dialects and opens the door to exploring the similarities between these dialects and other Arabic dialects.

Linguistic similarity between Maghreb dialect and Algerian dialect

The linguistic similarities among the Maghreb dialects, specifically Algerian, Tunisian, and Moroccan, manifest most prominently in their lexical domain. These dialects trace their linguistic roots to MSA, forming the cornerstone of their everyday vocabulary. Nonetheless, despite the individual evolution of each dialect and the emergence of unique characteristics and expressions, they retain numerous commonalities that testify to their shared cultural and linguistic heritage.

In all three dialects, it can be recognized that a lot of words have their origins in Arabic,

even though they may sound a bit different when spoken. Additionally, the Maghreb dialects have been significantly influenced by the Berber language, evidenced by the prevalence of Berber loanwords and expressions in daily conversations. Furthermore, the dialects of Tunisia and Morocco, much like those of Algeria, have undergone the imprint of colonial languages, particularly in the lexicon pertaining to modern technology, education, and transportation.

For instance, in all three dialects, the term ‘ميكرو’ (computer) (pronounced with slight variations) can be traced back to the French language. Similarly, the words ‘ويكند’ (week-end), and ‘مترو’ (subway), among other borrowed terms. These examples underscore the lexical parallels among the Algerian, Tunisian, and Moroccan dialects, emphasizing the shared vocabulary that characterizes the Maghreb region.

At phonological level, the similarity between Maghreb dialects can be summarized in many points, included:

- The Maghrebi dialects use non-Arabic phonemes /g/, /p/ and /v/ which are mainly used in words borrowed from foreign languages beside the MSA phonemes [24, 130].
- For Algerian dialect, the (ط) is uttered as /dʰ/ (ض), whereas (ذ) and (ث) are mostly pronounced as /d/ (د) and /t/ (ت) like in Moroccan dialects [130].
- The letter (ق) is it is uttered as /q/ and /g/ for Algerian dialect like Tunisian and Moroccan dialects. In addition, the (ق) is uttered as the glottal stop /ʔ/ in west of Algeria like in Morocco dialect.
- The consonant (ج) can be pronounced /ǧ/ or /ʒ/ (for Tunisian dialect and in some Algeria cities).
- The glottal stop (Hamza) in Maghreb dialects becomes longer, disappears, or is replaced by a vowel or letter, like in other Arabic dialects [130].

At the morphological level, the Maghreb dialects exhibit striking similarities, particularly when contrasted with MSA. One of the most prominent distinctions lies in the absence of dual and feminine plural forms in the Maghrebi dialects. Consequently, plural nouns and adjectives share identical forms for both masculine and feminine genders, diverging from MSA, where distinct plural forms exist for masculine, feminine, and dual cases.

Furthermore, a noteworthy parallel emerges in verb conjugation patterns between the Tunisian dialect and the ANB dialect. Additionally, the Maghrebi dialects dispense with the singular word declension, encompassing the nominative, genitive, and accusative cases at the end of the word [130]. This contrasts with MSA, where noun and adjective endings change to reflect their grammatical function within the sentence.

Lastly, negation in the Maghrebi dialects is achieved by affixing the prefix ‘ما’ at the beginning and the suffix ‘ش’ at the end of the verb. For instance, in the Moroccan dialect, the verb ‘كول’ (to eat) is negated by forming ‘مانكولش’ (not eating), with the addition of ‘ما’ at the outset and ‘ش’ at the end [130].

When it comes to the syntactic level, the Maghrebi dialects share a high degree of similarity in the word order of the sentence. This commonality is significant because it greatly improves the learning process for language models. In the case of the Algerian, Tunisian, and Moroccan dialects, the word order of a declarative sentence is relatively flexible. However, the SVO order is the most commonly used form, as both SVO and VSO orders are present in these dialects [131–133].

The flexibility of word order in these dialects allows for greater variation in sentence structure, which is often used for emphasis or to convey different meanings. For example, in the Tunisian dialect, the sentence "He ate an apple" can be expressed as ‘هُوَ كَلَّ التَّفَاحَ’ (SVO) or ‘كَلَّ هُوَ التَّفَاحَ’ (VSO), both of which convey the same meaning but with a slightly different emphasis.

Furthermore, the use of particles is common in the Maghrebi dialects. Clitics can be used to indicate possession, negation, or emphasis, among other things. In addition, the Maghrebi dialects share a number of conjunctions and adverbs that are used to connect phrases and clauses in a sentence, such as ‘و’ and ‘بين’.

5.4.2.2 Method

Our proposed method revolves around knowledge transfer, starting from Arabic dialects to Maghrebi dialects, and subsequently from Maghrebi dialects to Algerian dialects. This method entails a comprehensive and systematic transfer of shared linguistic features that are inherent in these dialects. By doing so, we establish a framework that empowers us to harness this knowledge transfer in an incremental manner.

Figure 5.6 provides a visual representation of the architecture of our proposed method, which comprises two pivotal steps:

- The first step entails transferring knowledge from Arabic dialects to Maghreb dialects, designating Arabic dialects as the source domain and Maghreb dialects as the target domain. This step requires a detailed analysis of the linguistic features of Arabic dialects, identifying the commonalities and differences between them, and determining the specific features unique to the Maghreb dialects.
- The second step involves transferring knowledge from Maghreb dialects to Algerian dialects, where the source domain shifts to Maghreb dialects, while the target domain shifts to Algerian dialects. Similar to the preceding step, this stage requires a

thorough analysis of the linguistic characteristics of Maghreb dialects, alongside an understanding of the specific features that distinguish Algerian dialects.

To accomplish the knowledge transfer between the various dialects, we employ a transfer learning strategy. Specifically, within each step of our proposed strategy, we implement a transductive transfer learning approach, providing us access to both the source and target domains. In this configuration, our model undergoes training on the source domain and is subsequently retrained on the target domain, thereby enabling the transfer of knowledge between them.

The appropriate approach for this scenario is the instance-based transfer learning strategy, as we described in section 5.4.1. This approach aims to resolve differences in marginal distribution by adjusting the weights of samples in the source domain. These weighted samples are then utilized for training in the target domain, allowing the target learner to utilize only the relevant information from the source domain. As mentioned in section 5.2.2.1, Instance-based transfer learning is effective in transferring knowledge between dialects, as it can capture the subtle linguistic nuances and variations between dialects. By leveraging the similarities and differences between the various dialects, we can develop a more robust and accurate model for knowledge transfer.

This work aims to employ the specialization in transferring learning. and it is based on training a model sequentially on corpora by going through the different levels shown in

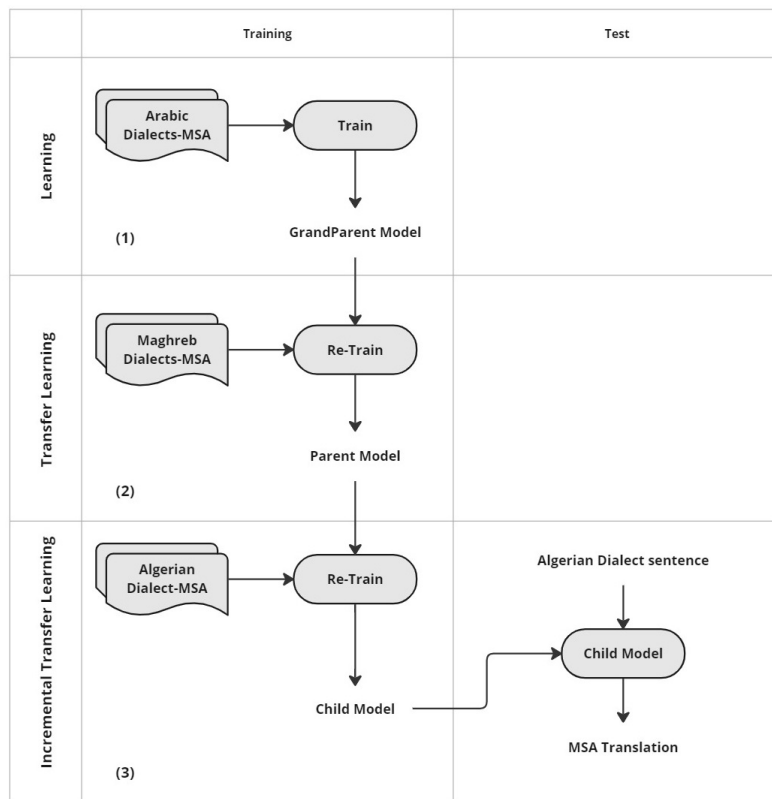


Figure 5.6: The Architecture of the Incremental transfer learning strategy used.

Figure 5.6. Formally, having a set of low parallel corpora $\begin{smallmatrix} x \\ y \end{smallmatrix} \Theta$ for language pairs (x, y) , the goal of our method is to maximize the sum of the log probability 5.8. Where, k is the number of NMT networks, θ_k is the generated parameters by NMT_k .

$$L_{\{\theta_k\}} = \sum_{k=1}^k \sum_{(a,b) \in \begin{smallmatrix} x \\ y \end{smallmatrix} \theta} \log p_k(\{b_j\}_{j=1}^m | \{a_i\}_{i=1}^n, \theta_k | \theta_{k-1}) \quad (5.8)$$

The proposed strategy entails the generation of a child model through the retraining of a parent model, which itself was initially pre-trained from a grandparent model. In this process, The parent model is responsible for providing training to the child model through a direct transfer of its knowledge. Simultaneously, the grandparent model contributes to the training of the parent model by directly sharing its knowledge. This approach allows for a gradual improvement of the model's performance by specializing it in the translation of a specific dialect.

By retraining the parent model with additional data from the target dialect, the child model is able to learn to translate with a higher degree of accuracy in that dialect. The knowledge transfer from the parent model to the child model allows for the child model to start with a strong foundation of pre-existing knowledge, and then refine that knowledge to specialize in the target dialect.

This approach is a form of transfer learning, which has been shown to be effective in improving the performance of NMT models by leveraging pre-existing knowledge. By gradually specializing the model in the target dialect, the resulting child model can achieve higher translation accuracy than the parent or grand-parent models alone. The strategy implementation involves the utilization of the attentional seq2seq and Transformer models that were presented in the preceding chapter.

5.4.2.3 Experimentation

To enhance the performance of NMT for the low-resource Algerian dialect, we employ an incremental transfer learning methodology. This approach involves a multi-step training process, starting with the training of attentional seq2seq and transformer models on a diverse set of Arabic dialects to create the grand-parent models. Subsequently, these grandparent models are further retrained on Maghrebi dialects to generate the parent models. Finally, the parent models are individually retrained on the Algerian dialect to produce the child models.

In addition to enhancing NMT for the Algerian dialect, we also sought to evaluate the effectiveness of this strategy and validate its utility. To do so, we expanded its application to address the challenges presented by low-resource Tunisian and Moroccan dialects,

demonstrating its adaptability and performance across multiple dialects.

To further improve results, we take into account the relationship between dialects. Since Maghrebi dialects share many similarities with various Arabic dialects, we leverage these shared characteristics to enhance the performance of the model trained on Maghrebi dialects. This specialized model is then retrained on Algerian, Tunisian, and Moroccan dialects separately to achieve better results for the target dialects.

While the grand-parent models are capable of translating all Arabic dialects, they may struggle to accurately translate rare words in the target dialect. By specializing the models on Maghrebi dialects through retraining, we significantly improve its performance in translating the target dialect. The trained child models are the most accurate in translating the target dialects, as it is specialized for them.

To train the grand-parent and child models, we adhere to the data partitioning described in Section 5. Specifically, Experiment 5-TL focuses on training for the Algerian dialect, Experiment 6-TL for the Tunisian dialect, and Experiment 7-TL for the Moroccan dialect. Additionally, we incorporate 23,650 sentences from the PADIC corpus to train the parent models. We label experiments employing this particular strategy as **Experiment 5-ITL** for the Algerian dialect, **Experiment 6-ITL** for the Tunisian dialect, and **Experiment 7-ITL** for the Moroccan dialect.

The utilized models are implemented in TensorFlow and Keras. Accuracy metrics and categorical cross-entropy loss are employed to evaluate the performance of the models during experimentation. The models are trained for 50 epochs with a batch size of 512 sequences per iteration. These parameters are not selected arbitrarily, but after conducting multiple experiments to achieve the best possible results. Prior to training, all sequences within a batch are padded to match the length of the longest sequence. We observe that all models exhibit convergence in terms of their training and validation accuracy, as well as their losses.

5.4.3 Result

This section aims to assess the effectiveness of the transfer learning strategies to improve the NMT model's performance by presenting experimental findings, detailed results explanations, BLEU metric scores, and sample translations.

In the previous chapter, experiments were conducted to evaluate seq2seq, attentional seq2seq, and Transformer models using the same corpora and models as the transfer learning experiments in this chapter. The previous experiments focused on translating Arabic dialects, while the current transfer learning experiments aim to showcase the effectiveness of using pre-trained models for improving low-resource Arabic dialect translation.

To gauge the transfer learning approach's effectiveness, the results from the experiments of the previous chapter are used as a benchmark for comparison with the transfer

learning experiment results in this chapter. This allows for a fair and objective evaluation of the transfer learning approach.

Score BLEU Result

Tables 5.2, 5.3, and 5.4 show the BLEU scores for three different scenarios: learning, transfer learning, and incremental transfer learning for three different Arabic dialects: Algerian, Tunisian, and Moroccan. Additionally, it also compares the performance of three different models: Seq2seq, Attentional Seq2seq and transformer.

For the Algerian dialect, we observed a significant improvement in BLEU scores. The seq2seq model increased the score from a mere 0.05 to a substantial 34.56, while the attentional seq2seq model elevated it from 16.5 to an impressive 35.78. The transformer model also displayed its prowess by boosting the score from 1.23 to 1.76. These results underscore the remarkable effectiveness of transfer learning in enhancing model performance.

Interestingly, the scores for the Tunisian and Moroccan dialects also exhibited noteworthy enhancements. With the attentional seq2seq model, the Tunisian dialect saw a remarkable rise from 0.03 to 1.23, and the Moroccan dialect followed suit, ascending from 0.04 to 1.27. Furthermore, the transformer model contributed significantly, pushing the Tunisian dialect from 1.26 to 2.14 and elevating the Moroccan dialect from 1.01 to an impressive 2.43. These improvements highlight the versatility of transfer learning strategy in bolstering the performance of models across different dialects.

Furthermore, the incremental transfer learning approach resulted in even higher scores. The Algerian, Tunisian, and Moroccan dialect scores increased significantly to 3.21, 1.75,

	Learning	Transfer Learning
Algerian Dialect	0.05	34.56

Table 5.2: BLEU scores results of conventional Transfer learning experiments evaluating the effectiveness of the seq2seq model in Algerian dialect translation. The previous chapter's learning experiment serves as a baseline to measure the impact of transfer learning.

	Learning	Transfer Learning	Incremental transfer learning
Algerian dialect (Experiment 2 (chapter 4))	16.5	35.87	-
Algerian dialect (Experiment 5 (chapter 4))	0.05	1.31	3.21
Tunisian dialect	0.03	1.23	1.75
Moroccan dialect	0.04	1.27	1.71

Table 5.3: BLEU scores results of incremental and conventional Transfer learning experiments evaluating the effectiveness of Attentional Seq2Seq model in Arabic dialects translation. The previous chapter's learning experiments serve as a baseline to measure the impact of transfer learning.

	Learning	Transfer Learning	Incremental transfer learning
Algerian dialect	1.23	1.76	6.52
Tunisian dialect	1.26	2.14	3.34
Moroccan dialect	1.01	2.43	2.54

Table 5.4: BLEU scores results of incremental and conventional Transfer learning experiments evaluating the effectiveness of the Transformer model in Arabic dialects translation. The previous chapter’s learning experiments serve as a baseline to measure the impact of transfer learning.

and 1.71 with the attentional seq2seq model, and 6.52, 3.34, and 2.54 with the transformer model, respectively.

These results highlight the potential of transfer learning, particularly incremental transfer learning, in improving the performance of NMT models in low-resource settings. By leveraging the available resources and knowledge, these models can be trained to accurately process and understand various dialects.

Translation Samples Results of Transfer Learning Experiment Using Seq2seq Model

In this part, we present the translation results obtained from the experiment conducted with the seq2seq model, incorporating learning, and transfer learning. The learning experiment conducted in the previous chapter serves as a baseline for evaluating the experiments introduced in this chapter (which is Experiment 2). Table 5.5 showcases the accurately predicted translations generated by the seq2seq model, while Table 5.6 presents the translations that are partially correct.

Tables 5.5 and 5.6 reveal that the seq2seq model completely failed in translating any sentence correctly in the learning experiment. furthermore, this model gave very different translations from the content of the source sentence, which is a major drawback.

Moving on to the transfer learning experiment, Table 5.5 shows that the seq2seq model managed to correctly translate some sentences. However, table 5.6 shows that the model still struggled with some sentences, and only partially accurately translated them.

With regard to partially correct sentences, the model, with transfer learning, provided translations that lacked some words or made mistakes in some words (such as in sentence 4), and this is what makes the meaning of the predicted sentence completely different from the original sentence.

Translation Samples Results of Transfer Learning Experiments Using Attentional Seq2seq Model

In this part, we present the translation results of various experiments, including learning, and transfer learning conducted with the Attentional seq2seq model. The learning ex-

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	1	أرواحي	تعالى	come here	نعم	yes	تعالى	come here
2	4	و أنت كاش مغامرات	و أنت هل من مغامرات؟	And you, are you an adventurer?	و أنت ما أحوالك؟	And how are you doing?	و أنت هل من مغامرات؟	And you, are you an adventurer?
3	6	راه مشوكي دوك سور كابتة حاجة	إيه مصدوم الآن أكيد أن ه ناك شينا	He is shocked now, surely there is something	أنت أيضا صرخت في وج هها	You also shouted in her face.	إيه مصدوم الآن أكيد أن ه ناك شينا	He is shocked now, surely there is something
4	9	توفي أبسط الحوايج و الله الوقت هدا ولي واع	أنظر لأبسط الأشياء صدق بلي أن هذا الزمن صار صعبا	Look at the simplest things, believe me, this time has become difficult	هذا الذي أمامك من الحزن نر ما الدنيا	What's the situation in Algeria, where you are from?	أنظر لأبسط الأشياء صدق بلي أن هذا الزمن صار صعبا	Look at the simplest things, believe me, this time has become difficult
5	21	قادر يكون ثالث يوم قادر يكون رابع يوم راهم هدر بالصح الصبح ما راهوش باين دوك ملما راهي تم دوك تفهم ال صبح	إحتمال أن يكون ثالث يوم و احتمال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أسي هناك سيقم الوضع	It may be the third day or it may be the fourth day, they talked to you but the truth is unclear. my mother there, she will understand the situation	لما رآته أمه قالت له يجب أن يذهب من هنا فلت له ثم أكل الكلب يمكن أن يه اجننا و لا يعلمون عليها ن تتخلص منها	When his mother saw what he was doing, she told him that he must leave. I told him that just like a dog that has eaten might attack us, and we don't know what it's capable of, she should get rid of it.	إحتمال أن يكون ثالث يوم و احتمال أن يكون رابع يوم لقد تحدثوا لكن الحقيقة غير واضحة أسي هناك سيقم الوضع	It may be the third day or it may be the fourth day, they talked to you but the truth is unclear. my mother there, she will understand the situation
6	26	حقار راني كونفيميت مع البروف على لكريد كي تقدي اعلاينا اقرو ميل على كريدني تاغ دا و لا تاغ حابوت يعني لي كايين كامل	على فكرة لقد أكتت أمرا لقرض العقاري مع الأستاذ د، و أنت عندما تقرر عين أعلنا مثلا عن قرض منزل ل أو محل ، يعني كل الحا لات الموجودة	By the way, I have confirmed the matter of the real estate loan with the professor, and when you have time, please give us an example of a home or shop loan, meaning all the available cases.	و هل ذهبا إلى الغابة و هم غير معتادين على الصيد ثربوا أن يحضروا لهم اللح م لا يعرفون كيف يصطاد ون	And did they go to the forest while being unaccustomed to hunting? They were raised to have meat prepared for them and don't know how to hunt.	على فكرة لقد أكتت أمرا لقرض العقاري مع الأستاذ د، و أنت عندما تقرر عين أعلنا مثلا عن قرض منزل ل أو محل ، يعني كل الحا لات الموجودة	By the way, I have confirmed the matter of the real estate loan with the professor, and when you have time, please give us an example of a home or shop loan, meaning all the available cases.

Table 5.5: Examples of Correct Translations Generated by Seq2Seq Model with Transfer Learning in Experiment 2-TL. The Table presents the translations generated by the Seq2Seq model after the transfer learning (Experiment 2-TL), compared to the translations generated in Experiment 2 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in transfer learning	English translation	Predicted sentence with transfer learning	English translation
1	2	الناس كامل	كل الناس	All the people	من أجل	in order to	الناس الناس	All the people
2	4	كيفاش أنا حططني ع شرة	و كيف أعطيتي عشرة	And how did she give me ten?	ما الذي تقصدين؟ كيف ذلك؟	What do you mean? How so?	كيف أعطيتي	And how did she give me ten?
3	5	دوك نهدر معاه على كاترين	سألتكم معاه الآن عن كا ترين	I will talk to him now about Catherine	لو كنت أختك لأعطيتي إياهم مجانا	If you were my sister, I would give them to you for free	سألتكم معاه الآن	I will talk to him now about Catherine
4	5	لا لا اليوم قرينا على لشرة	لا درسنا اليوم على الما شرة	No, we have a lesson today at ten o'clock.	لا ، لقد أعطوا فعلا	No, they actually gave	لا درسنا اليوم	No, we have a lesson today at ten o'clock.
5	6	ما كاين والو كاينة غ ير الخدمة	لا يوجد شيء يوجد العمل فقط	There is nothing, only work	لا أعرف إن كان ساعت ي مضبوطة	I don't know if my watch is accurate.	هذا لا يوجد شيء من إن شئت	There is nothing, only work
6	13	سليمة على لا تات تا عك شوفي هاد لي سقمونثور مور غولج يك مادا مور قناقر سا با	فيما يتعلق بأطروحتك سليمة ألقى تخطئة على أدوات التجزئة المورفولوجية ماذا مور قناقر سا با	Regarding your thesis, take a look at morphological segmentation tools. MorphTagger Saba?	أعلمين أن الدقائق الخمس الأولى من الكلمة سيظهر دائما مخصصة ل لوم لدرجة أنك تفتنين إنهاء الكلمة فوراً	Do you know that the first five minutes of a call with Suheir are always dedicated to blaming her to the extent that you wish to end the call immediately?	فيما يتعلق بأطروحتك سليمة عندما تكونين في المنزل ألقى البطارية أصلا واعلمي	Regarding your thesis, take a look at morphological segmentation tools. What is MorphTagger Saba?

Table 5.6: Examples of Partially Correct Translations Generated by Seq2Seq Model with Transfer Learning in Experiment 2-TL. The table presents the translations generated by the Seq2Seq model after the transfer learning (Experiment 2-TL), compared to the translations generated in Experiment 2 (which is without transfer learning) presented in the previous chapter.

periments from the previous chapter are used as a baseline to evaluate the experiments presented in this chapter (which are experiments 2,5, 6, and 7). The presented Tables are as follows:

- Tables 5.7, and 5.8 display the correct and partially correct predicted translations to evaluate the results from experiment 2-TL with attentional seq2seq.
- Table 5.9 and 5.10 display the correct and partially correct predicted translations to evaluate the results from experiment 5-TL.
- Table 5.11 and 5.12 display the correct and partially correct predicted translations to evaluate the results from experiment 6-TL.
- Table 5.13 and 5.14 display the correct and partially correct predicted translations to evaluate the results from experiment 7-TL.

From Table 5.7, the attentional seq2seq model achieved a much better performance in the learning experiment. It managed to translate six sentences out of twelve correctly (sentences 2, 3, 4, and 5 from table 5.7 and sentences 3 and 5 from table 5.8). This model was able to maintain the meaning of the source sentence while producing an accurate translation.

Moving on to the transfer learning experiment, Table 5.7 shows that attentional seq2seq model managed to correctly translate some sentences. However, Table 5.8 shows that the model still struggled with some sentences, and only partially translated them.

Finally, we observe that some sentences were predicted correctly in the learning experiment with the attentional seq2seq model but were predicted incorrectly in the transfer learning phase with the same model. For instance, sentences 3 and 5 from Table 5.8 were predicted correctly in the learning experiment, but they were partially translated in the transfer learning experiment. This observation indicates that the performance of the attentional seq2seq model may depend on the type of sentence being translated and that the model may require further fine-tuning to improve its accuracy.

After examining the results from both the learning and transfer learning experiments, it is evident that the attentional seq2seq model struggled to achieve high accuracy in translating the Algerian dialect to MSA. From Tables 5.9 and 5.10, we observe that the model did not succeed in translating most of the sentences, but it did manage to produce four partially correct translations out of twelve sentences in the learning experiment.

However, when we conducted the transfer learning experiment, we noticed a marked improvement in the model's performance. The transfer learning model was able to learn from its previous experience and utilize that knowledge to enhance its translation accuracy. This was evident in the correct translations that we observed in table 5.9, as well as the partially correct translations in Table 5.10.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with learning	English translation
1	1	أرواحي	تعالى	Come here.	يا الهي	Oh my god	تعالى	Come here.
2	3	واش راكي ديوي	مدا تظلين؟	What are you doing?	ومدا تظلين؟	What are you doing?	ومدا تظلين؟	What are you doing?
3	6	كي ابتي تكرر هي حنا وا قش نديرو	أنت سلمات؟ ههنا نفعل نج ن؟	Are you tired? What should we do?	أنت سلمات؟ ههنا نفعل نج ن؟	Are you tired? What should we do?	أنت سلمات؟ ههنا نفعل نج ن؟	Are you tired? What should we do?
4	11	قال لي اخنت شمام واشر راكي فاك له لاياش ال حمد الله	قال لي اخنت شمام كيف ح الك فاك له بخير و الحمد الله	He asked me, the sister of Shamaam, how I was doing, and I replied that I'm fine, thanks be to God.	قال لي اخنت شمام كيف ح الك فاك له بخير و الحمد الله	He asked me, the sister of Shamaam, how I was doing, and I replied that I'm fine, thanks be to God.	قال لي اخنت شمام كيف ح الك فاك له بخير و الحمد الله	He asked me, the sister of Shamaam, how I was doing, and I replied that I'm fine, thanks be to God.
5	16	حتي انا دايمن في الرو تئين شفتيت على الجمعة لي فاك جازت حلوة و الجمعة هدي قاع	أنا أيضا دائما في الروتين ، أفكر أن الجمعة الماضية مرت رائعة أما هذه الجمع ة فسيئة جدا	I'm also always in the routine, I remember that last Friday was great but this Friday is very bad	أنا أيضا كنت في الروتين ، و الحمد لله اذهب و ترو ج إلى الرياضيات و لا أدع ب	I was also in the routine, thank God go and marry to mathematics and not go	أنا أيضا دائما في الروتين ، أفكر أن الجمعة الماضية مرت رائعة أما هذه الجمع ة فسيئة جدا	I'm also always in the routine, I remember that last Friday was great but this Friday is very bad
6	27	حقا راكي كونفيميت م ع البروف على الكريد ي تاغ ليوميليني و انت كي تقدي اعطينا اقرو ميل على كريدني تاغ دا ر ولا تاغ جايوت يعني لي كايين كامل	على فكرة لقد أكدت أمر ا لقرض العقاري مع الأستاذ ، و أنت عندما تقرعين أعل نا مثلا عن قرض منزل أ و محل ، يعني أن هناك الم وجودة	By the way, I confirmed the mortgage issue with the professor, and when you have time, give us an example of a home or shop loan, meaning all possible scenarios.	على فكرة لقد أكدت أمر ا لقرض العقاري مع الأستاذ ، و أنت عندما تقرعين أعل نا مثلا عن قرض منزل أ و محل ، يعني أن هناك الم وجودة	By the way, I also confirmed the mortgage with the expert. When you have time, please give us an example of a home or store loan, meaning all possible cases.	على فكرة لقد أكدت أمر ا لقرض العقاري مع الأستاذ ، و أنت عندما تقرعين أعل نا مثلا عن قرض منزل أ و محل ، يعني أن هناك الم وجودة	By the way, I confirmed the mortgage issue with the professor, and when you have time, give us an example of a home or shop loan, meaning all possible scenarios.

Table 5.7: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 2-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 2-TL), compared to the translations generated in Experiment 2 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	1	بزاز	كثيرا	A lot	من التعب	because of tiredness	الكثير	A lot
2	3	الحمد لله فرات	الحمد لله واخيرا	finally, Praise be to God	الحمد لله	Praise be to God	الحمد لله وانت	finally, Praise be to God
3	8	والله غير تشوف يا ايمان ن يا اختي	صدقني يا اختي ايمان ان ي ساري	Believe me, my sister Iman, I will see	صدقني يا اختي ايمان في ط	Believe me, my sister, Iman only	صدقني يا سارة صدقني اني ساري	Believe me, my sister Iman, I will see
4	10	راهي ولاات الأزمة نار تو ، مام اورويا ما ولا تقن تصلح الهجرة	اصبحت الأزمة في كل م كان حتى اوريا لم تدت صلح الهجرة	The crisis has become everywhere, even Europe is no longer suitable for immigration.	اشعلت المدقة ولم تنتبه بد ت الحقيقة التي يحملها سوب بالاشتمال	She lit the heater and didn't notice the bag carrying the computer started to catch fire	اصبحت الأزمة في مكان ما باثقة أو تيسة و لم يكن غالي الثمن	The crisis has become everywhere, even Europe is no longer suitable for immigration.
5	12	لا لا انا نمرف وحدة را حت لفرنسا و عندها ب يت حاملة فيها فراتها	لا اعرف واحدة ذهبت إلى فرنسا و لها غرفة و صنع ت فيها الأفرشة	I don't know anyone who went to France and had a room with mattresses in it.	لا اعرف واحدة ذهبت إلى فرنسا و لها غرفة و صنع ت فيها الأفرشة	I don't know anyone who went to France and had a room with mattresses in it	لا اعرف واحدة ذهبت إلى فرنسا و لها مائدة لنحل خالي و نحضر الجورت و نحضر الفروض و نحضر الامتدادات	I don't know anyone who went to France and had a room with mattresses in it.
6	25	على بالك العام اللي فا ت قال لي واحد داوين تشوفه حاب بينال القسم و تشوفه داوين في اللط رينق ما على باليتن بلي هناك هو الولد	أتعلمين ؟ العام الماضي قا ل لي شخص أراه دائما ف ي الطريق أنه يريد أن ي ضع ولده في قسم آخر ، كنت أراه دائما في الطريق ولم أكن أعلم أنه هو والد الطفل	Do you know? Last year, a person who I always see on the street told me that he wants to put his son in another department. I always saw him on the street and didn't know he was the child's father.	هذه كذالك كنت في البداية ن كنت صدقي في المقرو ض أن أنه كنت في المقروض أن أسمع أحدهم يوم كنت كنت و كنت أعلم أنا أحب	This is how I was in the beginning if you believed me in that I was supposed to hear someone the day I was and I knew I love	أتعلمين ؟ العام الماضي قا ل لي شخص أراه دائما ف ي الطريق أنه يريد أن ي ضع ولده في قسم آخر ، كنت أراه دائما في الطريق	Do you know? Last year, a person who I always see on the street told me that he wants to put his son in another department. I always saw him on the street and didn't know he was the child's father.

Table 5.8: Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 2-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 2-TL), compared to the translations generated in Experiment 2 (which is without transfer learning) presented in the previous chapter.

Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	وين القهوة؟	أين المقهى؟	Where is the café?	أين؟	where?	أين المقهى؟	Where is the café?
2	تسكن من؟	هل تعيش بالقرب من هنا؟	Do you live near here?	ثمانية عشر	Eighteen	هل تعيش بالقرب من هنا؟	Do you live near here?
3	نخلص خمسين دولار	أنا سوف أدفع خمسون دولاراً	I will pay fifty dollars	والفضل يعود لصباح	Thanks to Sabah	أنا سوف أدفع خمسون دولاراً	I will pay fifty dollars
4	نص قرعة، من فضل لك	نصف زجاجة من فضل لك	Half a bottle, please	في اليسار، من فضلك	To the left, please	نصف زجاجة من فضل لك	Half a bottle, please
5	شحال في عمرها بن تاك؟	كم عمر ابنتك؟	How old is your daughter?	كم أكره هذا الكلام	How much I hate this talk	كم عمر ابنتك؟	How old is your daughter?
6	رائي حاب حديد تاغ سفر	أريد مكواة سفر	I want a travel iron	أريد أن أرى هذا	I want to see this	أريد مكواة سفر	I want a travel iron

Table 5.9: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 5-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 5-TL), compared to the translations generated in Experiment 5 (which is without transfer learning) presented in the previous chapter.

	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	4	صباح الخير واش راك؟	صباح الخير كيف حالك	Good morning, how are you?	مرحبا كيف حالك؟	Hello, how are you?	مرحبا كيف حالك؟	Hello, how are you?
2	4	ماكاش صابون في غرفتي	ليس هناك صابون في حجرتي	There is no soap in my room	لا يوجد هناك	There is nothing here	ليس هناك صابون في الليل	There is no soap at night
3	7	جامي سمعت بهاذ العنوان في هاذ الجبهة	لم اسمع بهذا العنوان من هنا	I haven't heard of this address before near here.	لم تقولوا شيئا حول العنوان الذي أرسلته	You didn't say anything about the article that I sent	لم اسمع العنوان من هنا	I did not hear the address from here
4	8	راني متأكد بلي درتو فالحبيب الدخاني تم ديالي	أنا متأكد أنني وضعتها في الحقيبة الداخلية للمعطف	I am sure that I put it in the inner pocket of the coat.	جعلت هذا القميص صديقا أرسلته	I made this shirt specifically for sending	أنا متأكد أنني وضعتها في محفظة الأوراق الخاصة بي	I'm sure that I put it in my wallet
5	8	هل شركة الطيران تخلص لي شراء الأعراس الشخصية؟	هل ستدفع لي شركة الطيران لشراء بعض الاحتياجات الشخصية الأساسية؟	Will the airline pay for me to purchase some basic personal needs?	إن الرصيف رقم خمسة يقع هناك	Pier five is located there	هل ستدفع لي شركة الطيران لشراء بعض الاحتياجات؟	Will the airline pay for my purchase of some necessities?
6	10	يزي برك في الصباح وما قرينتيش تقرا في الليل	هذا يكفي لم تدرس في الصباح حتى تدرس في الليل	That's enough. You didn't study in the morning until you studied at night	إننا أردت أن يكون هذا أيضا كل شيء؟	If you want this to be too that's it?	قلت لها في الجامعة أنني لم تدرس بعد في الجامعة	I told her at the university even if she has not yet studied at the university

Table 5.10: Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 5-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 5-TL), compared to the translations generated in Experiment 5 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in transfer learning	English translation	Predicted sentence with transfer learning	English translation
1	2	وينو المصعد؟	أين المصعد؟	Where is the elevator?	أين الإثناء؟	Where is the bowl?	أين المصعد؟	Where is the elevator?
2	4	تتجم تبيلهم للولار. يعبيتك؟	هل لك أن تغيرهم إلى د ولارات من فضلك؟	Can you change them into dollars, please?	هل يمكن أن تقول لي ك يف يمكن تحديله؟	Can you tell me how it can be adjusted?	هل لك أن تغيرهم إلى د ولارات من فضلك؟	Can you change them into dollars, please?
3	4	شنوة الخبز الي عند كم؟	ما نوع الخبز لديكم؟	What type of bread do you have?	هناك من هناك القليل	There from there they like	ما نوع الخبز لديكم؟	What type of bread do you have?
4	4	قداش وقت من هوني؟	كم المدة التي يستغرقها من هنا؟	How long does it take from here?	كم من محطة حتى يمكني التغير؟	How many stations until I can make a transfer?	كم المدة التي يستغرقها من هنا؟	How long does it take from here?
5	5	مئين النجم نشر ي زر بية متاع بحر؟	أين يمكنني أن أشتري حصيرة للشاطئ؟	Where can I buy a beach mat?	هل يمكنني أن أرسل بعض اللوازم في معدتي؟	Can I send some medicine for my stomach pain?	أين يمكنني أن أشتري حصيرة للشاطئ؟	Where can I buy a beach mat?
6	8	انجم نقص زور تساكرك درجة ثانية لشيكافغو، يعبيتك؟	هل يمكنني الحصول على تذكرتين درجة ثانية لشيكافغو من فضلك؟	Can I get two second-class tickets to Chicago, please?	هل تستطيع أن تأخذ ملابسك للوبي صباح الغد؟	Can you take my clothes to the lobby tomorrow morning?	هل يمكنني الحصول على تذكرتين درجة ثانية لشيكافغو من فضلك؟	Can I get two second-class tickets to Chicago, please?

Table 5.11: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 6-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 6-TL), compared to the translations generated in Experiment 6 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	2	مقيت لنويورك	لقد ذهبت إلى نيويورك	I went to New York	ذهبت إلى الجامعة	I went to the university.	ذهبت إلى الجامعة	Can I go to this job, please?
2	3	صباح الخير شنجوال ك؟	صباح الخير كيف حالك	Good morning, how are you?	صباح الخير كيف الحال	Good morning, how are you?	مرحبا كيف حالك؟	Hello, how are you?
3	3	تحب تبيت تحت الك بوت؟	هل تريد أن افحص موتور السيارة؟	Do you want me to check the car engine?	هل هناك أماكن شاعرة في رحلة اللدن؟	Is there any availability in the London trip?	هل تريد أن افحص موتور السيارة؟	Do you want me to check the car engine?
4	4	بانش ندفع خمسين دولار	أنا سوف أدفع خمسون دولاراً	I will pay fifty dollars	لقد أتناول كأساً من النبيذ	Have a glass of wine	أنا سوف خمسين دولار	I'll fifty dollars
5	6	الاستقبال؟ فماتى م صحة قريبة من هون ي؟	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a nearby clinic?	هل تمة مطعم قريب من هنا؟ هل مطعم صيني من هنا؟	Is there a restaurant nearby, a Chinese restaurant here?	هل هذا مكتب الاستقبال؟ هل تمة؟	Is this reception desk is there?
6	7	وقت توصل اطلب من واحد آخر، يبيتك	عندما تصل هناك أسأل شخص آخر، من فضلك	When you arrive there, please ask another person, please	هل يمكن أن أذهب إلى ه ذا العمل و من فضلك؟	Can I go to this job, please?	عندما تجيبها قولي لي الكلام و هناك من فضلك	When you find her, tell me the word and there, please

Table 5.12: Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 6-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 6-TL), compared to the translations generated in Experiment 6 (which is without transfer learning) presented in the previous chapter.

Upon analyzing Tables 5.11 and 5.12, we observe that the attentional seq2seq model was only able to successfully translate one sentence out of twelve sentences in the learning experiment. The sentence ‘صباح الخير كيف حالك’ was translated to ‘صباح الخير كيف الحال’, where the predicted translation omitted the letter ‘ك’, thus slightly altering the meaning of the question. However, the model partially translated three out of the twelve sentences correctly in the learning experiment.

On the contrary, our transfer learning experiment produced a notable enhancement in the model’s performance. We observed precisely translated sentences in Table 5.11 and partially accurate translations in Table 5.12. This improvement in performance can be attributed to the utilization of pre-existing knowledge from a different corpus, which allowed the model to learn from a wider range of linguistic features.

Upon analyzing Tables 5.13 and 5.14, we observe that the attentional seq2seq model did not succeed in translating any sentence accurately in its learning phase. However, it did manage to translate five out of twelve sentences partially correctly. Interestingly, one of the translated sentences was ‘صباح الخير كمال كيف حالك’ to ‘صباح الخير كيف حالك’, where the predicted translation included the name "Kamal" which was not present in the original translation. Despite this error, the translated sentence still maintained the correct context.

Nonetheless, in the transfer learning phase, we notice that the model’s performance improved significantly. This is evident by the correct translations of sentences in Table 5.13 and partially correct translations in Table 5.14. The model has demonstrated its ability to generalize and apply what it has learned from the training phase to the new data in the transfer learning phase.

Translation Samples Results of Transfer Learning Experiments Using Transformer Model

In this part, we present the translation results of various experiments, including learning, and transfer learning conducted with the Transformer model. The learning experiments from the previous chapter are used as a baseline to evaluate the experiments presented in this chapter (which are experiments 5, 6, and 7). The presented tables are as follows:

- Table 5.15 and 5.16 display the correct and partially correct predicted translations to evaluate the results from experiment 5-TL.
- Table 5.17 and 5.18 display the correct and partially correct predicted translations to evaluate the results from experiment 6-TL.
- Table 5.19 and 5.20 display the correct and partially correct predicted translations to evaluate the results from experiment 7-TL.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	4	صباح الخير كيف دا يبر؟	صباح الخير كيف حالك	Good morning, how are you?	صباح الخير كمال كيف حالك	Good morning, how are you?	صباح الخير كيف حالك	Good morning, how are you?
2	5	واش ممكن نكري مع دات الخطس؟	هل أستطيع تأجير أداة للتنفس أثناء السباحة؟	Can I rent a breathing apparatus while swimming?	هل يمكنني استئجار مضرب وكرات؟	Can I borrow a racket and balls?	هل أستطيع تأجير أداة للتنفس أثناء السباحة؟	Can I rent a breathing apparatus while swimming?
3	5	عندي تقدير كبير للموسيقى الكلاسيكية	لدي تقدير كبير للموسيقى الكلاسيكية	I have a great appreciation for classical music	لدي موهبة فطرية	I have a natural talent	لدي تقدير كبير للموسيقى الكلاسيكية	I have a great appreciation for classical music
4	5	شحال كايصوي البريد الجوي لليابان؟	ما تكلفة البريد الجوي لي اليابان؟	What is the cost of air mail to Japan?	كم يبلغ من فضلك التمنينة	What is the price, please?	ما تكلفة البريد الجوي لي اليابان؟	What is the cost of air mail to Japan?
5	8	ميكي هايكاوا هادي هي الغرفة سبيع مية وجوج	ميكي هايكاوا هذه غرفة سبعة صفر اثنين	Miki Hayakawa This is room seven zero two	أجل الرحلة الجوية أربعة خمسة خمسة تسعة خمسة	Flight order four five five	ميكي هايكاوا هذه غرفة سبعة صفر اثنين	Miki Hayakawa This is room seven zero two
6	10	جوج صفر جوج تسعة ثلاثة تسعة تسعة صفر	اثنان صفر اثنان تسعة ثلاثة تسعة تسعة صفر	Two zero two nine three nine six seven zero	امش على طول النهر لمسافة نصف ميل تجد الحديقة	Walk along the river for half a mile and you will find the park	اثنان صفر اثنان تسعة ثلاثة تسعة تسعة صفر	Two zero two nine three nine six seven zero

Table 5.13: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 7-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 7-TL), compared to the translations generated in Experiment 7 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted sentence in learning	English translation	Predicted sentence with transfer learning	English translation
1	3	عافاك بلي الين لجن يهايت	من فضلك غير هذه البيا ت إلى جنهات	Please convert these pounds to Egyptian pounds.	من فضلك أعطني خمسين	Please give me fifty	من فضلك غير هذه من هذا الفندق	Please change this yen from this hotel
2	3	واحد لنويورك، عافاك	واحد لنويورك، من فضلك	One to New York, please.	لقد سقطت من قبل الطريق من قبل عشر	I fell by the way by ten	واحد لنويورك، لو سمحت	One for New York, please
3	4	فين كاين أقرب جزائر؟	أين أقرب محل جزائر؟	Where is the nearest butcher shop?	أين يمكنني أن أقيم أثناء الجولة؟	Where can I stay during the tour?	أين أقرب صيدلية؟	Where is the nearest pharmacy?
4	8	ممكن تعطيني جوج تذكرات درجة ثانية لشيكاجو، عافاك؟	هل يمكنني الحصول على تذكرتين درجة ثانية لشيكاجو من فضلك؟	Can I please get two second-class tickets to Chicago?	هل يمكنني أن أطلع على شهادة التطعيم الخاصة بك من فضلك؟	Can I see your vaccination certificate please?	قدر احصل على تذكرتين درجة سياحية لشيكاجو، لو سمحت؟	Can I get two economy class tickets to Chicago, please?
5	9	واشن هادا مكتب الا ستيقال؟ واشن كاين ستي عيادة قريبة؟	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a nearby clinic?	هل لك أي صناديق بريد قريبة من فضلك؟	Do you have any nearby mailboxes, please?	هذا مكتب الاستقبال؟ فيه عيادة قريبة؟	Is this the reception desk? Is there a nearby clinic?
6	10	أنا ماشي متأكد واشن حزممت حوايج لرحلة ديال ثلاثة أسابيع	أنا غير متأكد إذا كنت قد حزممت ملابس كافية لرحلة مدتها ثلاثة أيام	I'm not sure if I've packed enough clothes for a three-day trip	لا أفهم يخبرني لقمتي أنترعها من قم السبع	I don't understand, I am baffled by my morsel, which I snatch from the lion's mouth	أنا أنا متأكد أنني إذا املأيت كافية لرحلة مدتها أنا أنا جدا	I'm pretty sure I'll get enough clothes for the trip I'm on too

Table 5.14: Examples of Partially Correct Translations Generated by Attentional Seq2Seq Model with Transfer Learning in Experiment 7-TL. The table presents the translations generated by the Attentional Seq2Seq model after the transfer learning (Experiment 7-TL), compared to the translations generated in Experiment 7 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	1	أه	نعم	Yes.	نعم	Yes.	نعم	Yes.
2	2	أه صح	نعم صحيح	Yes, that's correct.	نعم صحيح	Yes, that's correct.	نعم صحيح	Yes, that's correct.
3	3	لطوكيو، من فضلك	إلى طوكيو، من فضلك	To Tokyo, please.	من أجل فضلك	For please	إلى طوكيو، من فضلك	To Tokyo, please.
4	3	واش راح تفواك	مانا ساقول لك؟	What should I say to you?	مانا ستنقل له	What will you do for him?	مانا ساقول لك؟	What should I say to you?
5	3	سايها الحمد لله	بخير الحمد لله	I'm fine, thanks be to God.	لائس الحمد لله	No problem, thanks be to God.	بخير الحمد لله	I'm fine, thanks be to God.
6	4	ايه باي باي صحا	نعم مع السلامة	Yes, goodbye.	حسنا مع السلامة إذا	Okay, goodbye then.	نعم مع السلامة	Yes, goodbye.

Table 5.15: Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 5-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 5-TL), compared to the translations generated in Experiment 5 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	2	ما زال ما بيدناش	لم نبدأ بعد	We haven't started yet.	لم ينشر بعض جات	Some of them haven't been published	ما زلت لم أبدا معها	I still haven't started with her.
2	3	أموي سبيل في بلادنا	مستحيل في بلادنا	It's impossible in our country.	هيا خذي في هدا شيء	Come, take this thing.	بلادنا في بلادنا الكثير من	country in our country, there are many.
3	3	قوليلي لوسات زينب	ظك في السبع يا زين ب	I said in the seventh, Zainab.	نعم السابغ الآخر	Yes, the other seventh.	سوف أطلب منك	I will ask you.
4	6	علاالك ما فهمتش أنا و اتش راه صاير	أعلمين لم أفهم الآن لماذا	You know, I don't understand why now.	أعلمين ما ؟ لا أفهم م اذا أنه صعب إذا ؟	Do you know what? I don't understand what, is it difficult then?	أعلمين لم أفهم لما أصبح ت ماذا المنزل ل	Do you know what? I don't understand why it became what the house
5	6	و اليوم صالا واتش رحتي درتي	و من أجل ماذا ذهبت اليوم؟	And why did you go today?	و اليوم ماذا اليوم ماذا فعلت	And today, what about today? What did you do?	و ماذا عن ال مسلسل	And what about the serie?
6	9	دوك تقول لك حاجة خالتي زينب راهي هنا دوك	ماتقول لك شيئا ، خالتي زينب هنا الآن	I'll tell you something, my aunt Zainab is here now.	سأري نبييل و أمر مباشرة	I will see Nabil and give him a direct order.	أرني يدك أمس كتك	Show me your hand, I hold it

Table 5.16: Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 5-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 5-TL), compared to the translations generated in Experiment 5 (which is without transfer learning) presented in the previous chapter.

After analyzing Tables 5.15 and 5.16, it is observed that the transformer model was able to accurately translate only one sentence during its learning phase (sentence 1 in Table 5.15). However, it managed to meaningfully translate two sentences (sentences 5 and 6 in Table 5.15).

During the transfer learning phase, the model's performance showed significant improvement, as evidenced by correct translations in Table 5.15 and partially correct translations in Table 5.16, such as sentences 1 and 4. This illustrates how the model was able to apply its learning to Algerian dialect sentences, benefiting from the knowledge transferred from Arabic dialects to the Algerian dialect.

Upon examining Tables 5.17 and 5.18, it was noticed that the transformer model achieved precise translations for two sentences during its learning phase (sentences 1 and 2 in Table 5.17). Nevertheless, it was only able to translate the remaining sentences partially, with an average accuracy of only one to two words.

However, during the transfer learning phase, the model exhibited a remarkable improvement in performance. This was evident in the accurate translations displayed in Table 5.17 and partially correct translations in Table 5.18, for Tunisian dialect sentences. This highlights how the model leveraged its learning from Arabic dialects and applied it to the Tunisian dialect, leading to improved outcomes.

After reviewing Tables 5.19 and 5.20, it was observed that the transformer model delivered accurate translations for only three sentences during its learning phase (sentences 1 and 6 from Table 5.19, and sentence 2 from Table 5.20). However, for the remaining sentences, the model struggled and could only produce incorrect or partially correct translations, with an average accuracy of only one to two words.

However, in the transfer learning phase, the model demonstrated a significant enhancement in performance. This was evident in the precise translations presented in Table 5.19 and partially correct translations in Table 5.20, particularly for the Moroccan dialect sentences. This underscores how the model utilized its knowledge from Arabic dialects and applied it to the Moroccan dialect, resulting in better outcomes.

Translation Samples Results of Incremental Transfer Learning Experiments Using Attentional Seq2seq and Transformer Models

In this part, we present Tables 5.21, 5.22, 5.23, 5.24, 5.25, and 5.26 that demonstrate how the child-model in the incremental learning transfer experiment (experiments 5-ITL, 6-ITL, and 7-ITL) succeeded in translating the sentences that the parent model in the learning transfer experiments (experiments 5-TL, 6-TL, and 7-TL) could not translate completely correctly.

Tables 5.21, 5.22, 5.23, 5.24, 5.25, and 5.26 present a noteworthy finding about the effectiveness of the incremental transfer learning approach. These tables indicate that some

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	1	شَنو	ما ذا ؟	What?	ما ذا	What?	ما ذا ؟	What?
2	1	با هي	حسننا	Okay.	حسننا	Okay.	حسننا	Okay.
3	2	لطوكيو، يعني شك	إلى طوكيو ، من فضلك	To Tokyo, please.	غدا ، من فضلك	Tomorrow, please	إلى طوكيو ، من فضلك	To Tokyo, please.
4	2	وين المحطة ؟	أين المحطة ؟	Where is the station?	أين يمكنني أن آخذ هذا	Where can I take this?	أين المحطة ؟	Where is the station?
5	2	تلاتا أيام	ثلاثة أيام	Three days.	إنه أيام	It's days	ثلاثة أيام	Three days.
6	3	هذا كل شيء ؟	هل هذا كل شيء ؟	Is that everything?	هذا كل ما ك ان	That's all there was	هل هذا كل شيء ؟	Is that everything?

Table 5.17: Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 6-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 6-TL), compared to the translations generated in Experiment 6 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	2	من وقاش؟	منذ متى ؟	How long ?	من أجل أنها لم تقم	For she did not understand	من عندك	who is with you
2	3	ماينش خايبة منو	لست خائفة منه	I'm not afraid of him	لا خائفة منه	Not afraid of it	لست هذا مما يجب سل يمة	I am not this from what should correct
3	4	باهي ماتما حتى مشكل	حسنا لا مشكلة في ذلك	Well no problem with that	حسنا لم لا بأس في ذلك	Well that was okay	حسنا مشكلة ، هيا	Well problem, come on
4	4	فماشي متخيفة تككي يا باهي؟	هل هناك مخيفات يتحدثن اليابانية ؟	Are there any flight attendants who speak Japanese?	هل هل أنت جاد	Are you serious	هل هناك سيدي عاشور	Is there Sidi Ashour
5	5	باهي مالا مش مشكلة بسلامة	حسنا إذن لا مشكلة مع ال سلامة	Well then no problem bye	حسنا إذا لا يوجد أي شيء	Well if there is none	حسنا في أمن الله أعلم يا سارة	Well, in the safety of God knows best, Sarah
6	6	أول مرة نشوف وجهك كي هكا	أول مرة أرى وجهك بهذا الشكل	First time seeing your face like this	أول مرة عندما ناكل عندما يذهبون	The first time when we eat when they go	أول هذا كل هناك	First that's all there

Table 5.18: Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 6-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 6-TL), compared to the translations generated in Experiment 6 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	2	تمام بصح	نعم صحيح	Yes, that's correct.	نعم صحيح	Yes, that's correct.	نعم صحيح	Yes, that's correct.
2	3	ونت وانش كتقراي	و أنت هل تدرسين؟	And you, do you study?	و أنت هل تدرسين الآن سنذهب معك تدرسين ليلا	And you, are you studying now? We will go with you, study at night.	و أنت هل تدرسين؟	And you, do you study?
3	3	كيفاش تقدر فساعدك ؟	كيف استطيع مساعد تاك ؟	How can I help you?	كيف تقدر بمكنني أن أأخذ حقيبتي من هناك اتصل رجل يوالذي بخصوص البرامج الوطنية للبحث كيف سلطان الكباش	How can we help? Can I take my bag, please? Call my father's friend regarding the national research programs. How is Sultan Al-Kabash doing?	كيف استطيع مساعدك ؟	How can I help you?
4	4	بيعت مصلوح ديال ا لسفر	أريد مكواة سفر	I want a travel iron.	أريد مكواة سفر الأستاذ	I want a travel iron, Professor.	أريد مكواة سفر	I want a travel iron.
5	4	تماما غدي نبيط لها	تماما ساكسل بها	Certainly, I will call her.	تماما لينا بصدد كتابة الحروف الفرنسية	Exactly, we are not currently writing French letters.	تماما ساكسل بها	Certainly, I will call her.
6	5	واشن هادا مصنوع م ن القطن؟	هل هذا مصنوع من ا لقطن ؟	Is this made of cotton?	هل هذا مصنوع من القطن؟	Is this made of cotton?	هل هذا مصنوع من القطن ؟	Is this made of cotton?

Table 5.19: Examples of Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 7-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 7-TL), compared to the translations generated in Experiment 7 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English translation	Predicted translation in Learning	English translation	Predicted translation with Transfer learning	English translation
1	3	عشرة طواربع عداك	عشرة طواربع من فضلك	Ten stamps, please.	عشرة طواربع من فضلك فضلك	Ten stamps, please. Please. Please. I will take you to Canada. My daughter. My daughter. My daughter.	عشرة طواربع لو سمحت	Ten stamps, please
2	3	نعم، عندنا حجز ليك	نعم، لدينا حجز لك	Yes, we have a reservation for you.	نعم عندنا حجز لك	Yes, we have a reservation for you.	نعم ، عندنا حجز لك	Yes, we have a reservation for you.
3	3	نعم صباح الخير	نعم صباح الخير	Yes, good morning	هل هو طلب شخص الآن لا قمت لأن لا تكتفي للوالدي أمور كهذه لا اليوم ؟	Is it someone's request now? No, he didn't submit it now. Don't come to the governor for such matters, not today.	هل ذهبت إلى أم لا ؟	Did you go or not?
4	3	بنات خالي محمد	ابنة خالي محمد	My cousin's daughter, Mohammed.	نعم صباح الخير كيف صباح الخير	Yes, good morning. How good is the morning?	نعم صباح الخير كيف	Yes, good morning. How are you?
5	4	مزبان مكنين حتى مشكل	حسننا لا مشكلة في ذلك	Okay, no problem with that.	فاهة الشرف، أراقى العروس ... حصاتي و تقتف رجال الشرطة بالحجارة	Maid of honor, I accompany the bride... My mother-in-law and the police officers are throwing stones.	لقد أقل من المتصرف	It is less than expected.
6	5	واش بديكي دابا ولا لا ؟	هل بنات الآن أم لا ؟	Have you started now or not?	جيد لا نحن مشكلة لديها مشكلة لديها مشكلة لديها أطفال مشكلة لديها مشكلة لديها أطفال مشكلة لديها أطفال	Good, we don't have a problem. She has a problem. She has a problem. She has children, a problem. She has a problem. From problem to problem, she has children.	جيد هم الذين لم مشكلة	Good, they don't have a problem.

Table 5.20: Examples of Partially Correct Translations Generated by Transformer Model with Transfer Learning in Experiment 7-TL. The table presents the translations generated by the Transformer model after the transfer learning (Experiment 7-TL), compared to the translations generated in Experiment 7 (which is without transfer learning) presented in the previous chapter.

Number	Size	Dialect	Original translation	English Translation	Predicted translation	English Translation
1	4	صباح الخير واش راك؟	صباح الخير كيف حالك	Good morning, how are you?	صباح الخير كيف حالك	Good morning, how are you?
2	4	ماكانش صابون في غرفتي	ليس هناك صابون في حجرتي	There is no soap in my room	ليس هناك صابون في حجرتي	There is no soap in my room
3	7	جامعي سمعت بهذا العنوان في هاذ الجبهة	لم اسمع بهذا العنوان من قبل بال قرب من هنا	I haven't heard of this address before around here	لم اسمع بهذا العنوان من قبل بالقرب من هنا	I haven't heard of this address before around here
4	8	راني متأكد بلي درتو فالجيب الدخاني تح دياي	أنا متأكد أنني وضعتها بالجيب الداخلي للمعطف	I'm sure I put it in the inside pocket of the coat	أنا متأكد أنني وضعتها بالجيب الداخلي للمعطف	I'm sure I put it in the inside pocket of the coat
5	8	هل شركة الطيران تخلص لي شراء الأغراض الشخصية؟	هل ستدفع لي شركة الطيران لشراء بعض الاحتياجات الشخصية الأساسية؟	Will the airline pay for me to buy some basic personal necessities?	هل ستدفع لي شركة الطيران لشراء بعض الاحتياجات الشخصية الأساسية؟	Will the airline pay for me to buy some basic personal necessities?
6	10	يزي برك في الصباح وما في ريتيش تقرأ في الليل	هذا يكفي لم تدرس في الصباح حتى تدرسي في الليل	This is enough, you didn't study in the morning so that you can study at night	هذا يكفي لم تدرس في الصباح حتى تدرسي في الليل	This is enough, you didn't study in the morning so that you can study at night

Table 5.21: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Incremental Transfer Learning in Experiment 5-ITL. The table presents the translations generated by the Attentional Seq2Seq model after the incremental transfer learning, which are partially correct translations generated in Experiment 5-TL (which is with transfer learning).

Number	Size	Dialect	Original translation	English Translation	Predicted translation	English Translation
1	2	مقديت لنيويورك	لقد ذهبت إلى نيويورك	I went to New York	لقد ذهبت إلى نيويورك	I went to New York
2	3	صباح الخير شنوالك؟	صباح الخير كيف حالك	Good morning, how are you?	صباح الخير كيف حالك	Good morning, how are you?
3	3	تحب تبيت تحت الكوت؟	هل تريد أن افحص موتور السيارة؟	Do you want me to check the car engine?	هل تريد أن افحص موتور السيارة؟	Do you want me to check the car engine?
4	4	باش ندفع خمسين دولار	أنا سوف أدفع خمسون دولاراً	I will pay fifty dollars	أنا سوف أدفع خمسون دولاراً	I will pay fifty dollars
5	6	الاستقبال؟ فساتي مصححة قرينة من هوني؟	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a clinic nearby?	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a clinic nearby?
6	7	وكت توصل اطلب من واحد آخر، يمشك	عندما تصل هناك أسأل شخص آخر، من فضلك	When you get there, please ask someone else	عندما تصل هناك أسأل شخص آخر، من فضلك	When you get there, please ask someone else

Table 5.22: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Incremental Transfer Learning in Experiment 6-ITL. The table presents the translations generated by the Attentional Seq2Seq model after the incremental transfer learning, which are partially correct translations generated in Experiment 6-TL (which is with transfer learning).

Number	Size	Dialect	Original translation	English Translation	Predicted translation	English Translation
1	3	واحد لتيويورك، عافاك	واحد لتيويورك ، من فضلك	One to New York, please	واحد لتيويورك ، من فضلك	One to New York, please
2	4	عافاك بلى الين لحنيها	من فضلك غير هذه الينات إلى ج نيهات	Please convert these yen to pounds	من فضلك غير هذه الينات إلى جنيها	Please convert these yen to pounds
3	4	فين كاين أقرب جزار؟	أين أقرب محل جزار؟	Where is the nearest butcher shop?	أين أقرب محل جزار؟	Where is the nearest butcher shop?
4	8	ممكن تعطيني جوج تنكيرات درجة ثانية لشيكافو، عافاك؟	هل يمكنني الحصول على تنكرتين درجه ثانية لشيكافو من فضلك؟	Where is the nearest butcher shop?	هل يمكنني الحصول على تنكرتين درجه ثانية لشيكافو من فضلك؟	Where is the nearest butcher shop?
5	9	والش هذا مكتب الاستقبال؟ والش كاين شي عيادة قريبة؟	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a clinic nearby?	هل هذا مكتب الاستقبال؟ هل يوجد عيادة قريبة؟	Is this the reception desk? Is there a clinic nearby?
6	10	أنا مائتي متأكد والش جزمت حوايج لرحلة ديال ثلاثة أسابيع	أنا غير متأكد إذا كنت قد جزمت ملابس كافية لرحلة مدتها ثلاثة أيام	I'm not sure if I've packed enough clothes for a three-day trip	أنا غير متأكد إذا كنت قد جزمت ملابس كافية لرحلة مدتها ثلاثة أيام	I'm not sure if I've packed enough clothes for a three-day trip

Table 5.23: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Incremental Transfer Learning in Experiment 7-ITL. The table presents the translations generated by the Attentional Seq2Seq model after the incremental transfer learning, which are partially correct translations generated in Experiment 7-TL (which is with transfer learning).

Number	Size	Dialect	Original translation	Predicted Translation	English Translation
1	2	مزال ما بيدناش	لم نبدأ بعد	لم نبدأ بعد	We haven't started yet.
2	3	أميو سبيل في بلادنا	مستحيل في بلادنا	مستحيل في بلادنا	It's impossible in our country.
3	3	قوتلي لوساك زينب	ظك في السبع يا زينب	ظك في السبع يا زينب	I said in the seventh, Zainab.
4	6	علايك ما فهمتش أنا واش راه صاير	أعلمين لم أفهم الآن لماذا	أعلمين لم أفهم الآن لماذا	You know, I don't understand why now.
5	6	و اليوم مالا واش رحتي درتي	و من أجل ماذا ذهبت اليوم؟	و من أجل ماذا ذهبت اليوم؟	And why did you go today?
6	9	دوك تقول لك حاجة خالتي زينب راهي هنا دو ك	سأقول لك شيئاً ، خالتي زينب هنا الآن	سأقول لك شيئاً ، خالتي زينب هنا الآن	I'll tell you something, my aunt Zainab is here now.

Table 5.24: Examples of Correct Translations Generated by Transformer Model with Incremental Transfer Learning in Experiment 5-ITL. The table presents the translations generated by the Transformer model after the incremental transfer learning, which are partially correct translations generated in Experiment 5-TL (which is with transfer learning).

Number	Size	Dialect	Original translation	Predicted Translation	English Translation
1	2	من وقاش؟	منذ متى ؟	منذ متى ؟	How long ?
2	3	ماينش خايقة منو	لست خائقة منه	لست خائقة منه	I'm not afraid of him
3	4	باهي ماتما حتى مشكل	حسننا لا مشكلة في ذلك	حسننا لا مشكلة في ذلك	Well no problem with that
4	4	فماشي مخصبة تحكي باباني؟	هل هناك مخصبات يتحدثن اليابانية ؟	هل هناك مخصبات يتحدثن اليابانية ؟	Are there any flight attendants who speak Japanese?
5	5	أول مرة تنوف وجهك كي هكا	أول مرة أرى وجهك بهذا الشكل	أول مرة أرى وجهك بهذا الشكل	Well then no problem bye
6	5	باهي مالا متن مشكلة بسلامة	حسننا إذن لا مشكلة مع السلامة	حسننا إذن لا مشكلة مع السلامة	First time seeing your face like this

Table 5.25: Examples of Correct Translations Generated by Attentional Seq2Seq Model with Incremental Transfer Learning in Experiment 6-ITL. The table presents the translations generated by the Transformer model after the incremental transfer learning, which are partially correct translations generated in Experiment 6-TL (which is with transfer learning).

Number	Size	Dialect	Original translation	Predicted translation	English Translation
1	3	عشرة طوابع عافاك	عشرة طوابع من فضلك	عشرة طوابع من فضلك	Ten stamps, please.
2	3	نعم، عندنا حجز لراك	نعم، لدينا حجز لك	نعم، لدينا حجز لك	Yes, we have a reservation for you.
3	3	نعم صباح الخير	نعم صباح الخير	نعم صباح الخير	Yes, good morning
4	3	بنت خالي محمد	ابنة خالي محمد	ابنة خالي محمد	My cousin's daughter, Mohammed.
5	4	مزيان مكيين حتى مشكول	حسنًا لا مشكلة في ذلك	حسنًا لا مشكلة في ذلك	Okay, no problem with that.
6	5	واش بديتي دابا ولا لا	هل بدأت الآن أم لا؟	هل بدأت الآن أم لا؟	Have you started now or not?

Table 5.26: Examples of Correct Translations Generated by Transformer Model with Incremental Transfer Learning in Experiment 7-ITL. The table presents the translations generated by the Transformer model after the incremental transfer learning, which are partially correct translations generated in Experiment 7-TL (which is with transfer learning).

sentences, which were incorrectly or partially translated correctly in the transfer learning experiment, were translated correctly when incremental transfer learning was employed. This clearly demonstrates that the adopted approach significantly enhances the efficiency of the attentional seq2seq and the transformer models.

The improved translation accuracy can be attributed to the fact that the models are trained to transfer the common knowledge from Arabic dialects to the Maghrebi dialect. This initial training lays the foundation for specialized learning, where the model is then trained separately for Algerian, Tunisian, and Moroccan dialects. As a result, the models gain a better understanding of the nuances and subtleties of each dialect, thereby becoming more effective at translating them.

Moreover, the models' successful translation of these dialects highlights its ability to impart knowledge and learn from different dialects to improve its translation capabilities. This suggests that the models have not only learned well, but also effectively transferred knowledge across different dialects, leading to improved performance in translating them.

5.4.4 Discussion

Within this section, we delve into the impact of employing learning transfer in Arabic dialects NMT, examining three distinct models, seq2seq, attentional seq2seq, and transformer. The discussion is structured across several subsections, each exploring a different aspect. Initially, we analyze the consequences of applying Transductive learning transfer to Arabic dialects NMT. Subsequently, we investigate the impact of incremental transfer learning. Moreover, we delve into the impact of the attention mechanism on transfer learning NMT. Lastly, we explore the optimal sentence length for MT, aiming to identify specific sentence lengths that generate improved results in transfer learning NMT tasks.

The Impact of Transfer Learning on Maghrebi Dialects NMT

Our previously presented experiments focused on the use of transductive transfer learning strategy to improve the performance of NMT models for low-resource Maghrebi dialects. Specifically, we used three NMT models: seq2seq, attentional seq2seq and Transformer models.

To evaluate the effectiveness of the transfer learning strategy, we conducted a baseline experiment where we directly trained the NMT model on each specific low-resource dialect, including Algerian, Tunisian, and Moroccan dialects. We then retrained the same model using the transfer learning approach, where the model was first trained on the Arabic dialect and then retrained on the specific low-resource dialect.

After conducting the test phase, we obtained the blue score results, and the results showed that the transfer learning strategy significantly improved the performance of the models in translating the specific low-resource Maghrebi dialects as shown in Figures 5.7.

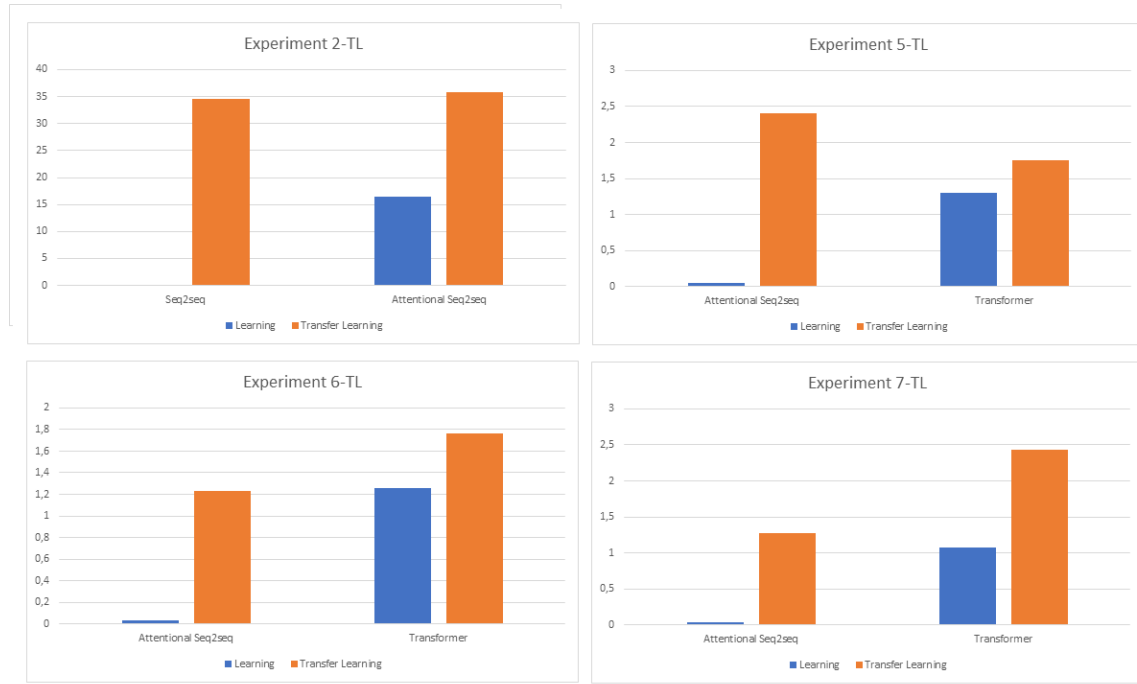


Figure 5.7: Unveiling the impact of Transductive Transfer Learning: A Comparative Analysis across Four Experiments with Seq2Seq, Attentional Seq2Seq, and Transformer Models.

For the Algerian dialect, the seq2seq model showed a dramatic improvement in the blue score, from 0.05 in the learning experiment to 34.56 in the transfer learning (Experiment 2-TL). The attentional seq2seq model also showed a significant improvement, with blue scores of 16.5 and 35.87 in the learning and transfer learning (Experiment 2-TL), respectively. Similarly, the attentional seq2seq model (Experiment 5-TL) showed a considerable improvement, with blue scores of 0.05 in the learning experiment and 1.31 in the transfer learning experiment. Furthermore, the transformer model experienced an increase in the blue score, rising from 1.23 in the learning experiment to 1.76 in the transfer learning (Experiment 5-TL).

For the Tunisian (experiment 6-TL) and Moroccan (experiment 7-TL) dialects, the attentional seq2seq model showed the most significant improvement, with blue scores increasing from 0.03 and 0.04 in the learning experiments to 1.23 and 1.27 in the transfer learning, respectively. Additionally, the Transformer model displayed the most remarkable enhancement, achieving substantial increases in blue scores. In the learning experiments, the blue scores were 1.26 and 1.01 for the Tunisian and Moroccan dialects, respectively. However, in the transfer learning experiments, these scores substantially improved to 2.14 and 2.43, respectively.

Our analysis indicated that the transfer learning strategy plays a crucial role in improving the accuracy of machine translation for Maghrebi dialects. The transferred knowledge from the Arabic dialect to the specific low-resource dialects helped the NMT models to

learn and generalize better, resulting in improved translations.

In summary, our results demonstrate that transductive transfer learning is a promising strategy for improving the performance of NMT models in low-resource settings. The use of transfer learning can help address the challenges of low-resource dialects and facilitate more accurate and effective machine translation across linguistic boundaries.

The Impact of Incremental Transfer Learning on Maghrebi Dialects NMT

Incremental transfer learning is a strategy where a model is initially trained on a particular language. Subsequently, the knowledge gained from this training is transferred to another model, which is then trained on under language. Finally, the knowledge acquired from the previous step is further transferred to another model, specifically trained on a low-resource specific under language. In this case, the objective is to use incremental transfer learning to benefit from the common knowledge between the Arabic dialects, the Maghrebi dialects, and the Algerian dialect to improve the performance of NMT models on Algerian dialect.

The proposed strategy involves starting with a grand-parent model that has been trained on the Arabic dialects. The knowledge from this model is then transferred to a parent model, which is trained on Maghrebi dialects. The parent model is specialized on the Maghrebi dialects by fine-tuning the model using data from this dialect. Next, the knowledge from the parent model is transferred to a child model, which is trained on Algerian dialect. The child model is further specialized on final target low resource dialect by fine-tuning the model using data from this dialect.

By using incremental transfer learning in this way, the child model benefits from the common knowledge between Arabic, Maghreb, and Algerian dialects. This strategy can potentially improve the performance of the model on the Algerian dialect task, as compared to training the model from scratch on Algerian dialect data. Furthermore, we applied this strategy on two under low resource dialects, which are Tunisian and Moroccan dialects.

Figures 5.8 shows that the incremental transfer learning strategy shown promising results in improving machine translation accuracy for low-resource dialects.

With attentional seq2seq model, for the Algerian dialect, the score bleu value was raised from 0.05 to 1.31 in learning and transfer learning (experiment 5-TL), respectively, and then further raised to 3.21 in incremental transfer learning (experiment 5-ITL). Similarly, for the Tunisian dialect, the score bleu value was raised from 0.03 to 1.23 in learning and transfer learning (experiment 6-TL), respectively, and then further improved to 1.75 with incremental transfer learning (experiment 6-ITL). Finally, for the Moroccan dialect, the score bleu value was raised from 0.04 to 1.27 in learning and transfer learning (experiment 7-TL), respectively, and then further improved to 1.71 through incremental transfer learning (experiment 7-ITL).

Moreover, utilizing the Transformer model, significant improvements were observed in

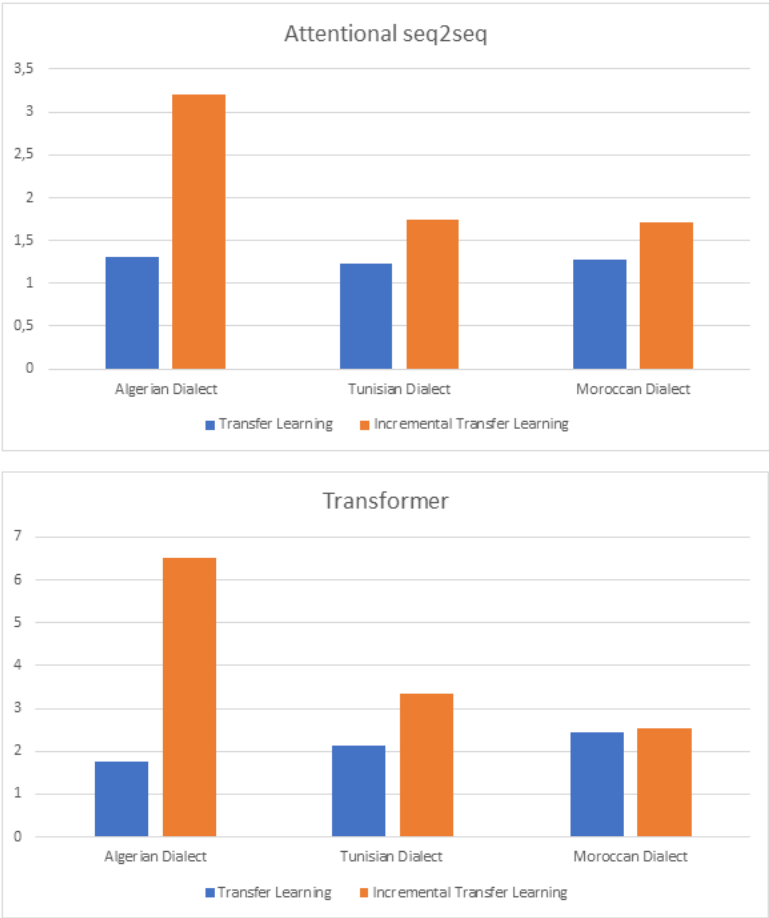


Figure 5.8: Unveiling the impact of Incremental Transfer Learning: A Comparative Analysis across three Experiments with Attentional Seq2Seq and Transformer Models.

the Algerian dialect. The bleu score increased from 1.23 to 2.14 in the learning phase and transfer learning (experiment 5-TL), respectively. Subsequently, with incremental transfer learning, the score further surged to 6.52 (experiment 5-ITL). Similarly, for the Tunisian dialect, the bleu score rose from 1.26 to 2.14 in the learning and transfer learning (experiment 6-TL) stages, respectively. Remarkably, through incremental transfer learning, the score further improved to 3.34 (experiment 6-ITL). Finally, concerning the Moroccan dialect, the bleu score increased from 1.01 to 2.43 during the learning and transfer learning phases (experiment 7-TL), respectively. Notably, incremental transfer learning resulted in an additional improvement, with the score reaching 2.54 (experiment 7-ITL).

Our analysis highlights the importance of incremental transfer learning in improving the accuracy of MT for Maghrebi dialects. The transferred knowledge from the Arabic dialect to the Maghreb dialect and then to the specific low-resource dialects helps the NMT models learn and generalize better, resulting in improved translations.

Furthermore, we note that the common knowledge imparted by specialization plays a crucial role in raising the efficiency of the model to translate specific under low-resource dialects. By leveraging the shared features of related dialects, incremental transfer learning can significantly improve the performance of NMT models on low-resource dialects.

In summary, incremental transfer learning has shown great potential in improving the accuracy of machine translation for low-resource dialects. By transferring knowledge from related dialects, NMT models can learn and generalize better, resulting in improved translations.

The Impact of Attention Mechanism of Transfer Learning NMT for Arabic dialects

In our previous experiment, we delved into the world of MT by utilizing two different models: seq2seq and attentional seq2seq (experiment 2-TL). The main goal of this experiment was to determine the effectiveness of the attention mechanism when it came to transfer learning.

Our experiment was focused solely on the Algerian dialect, as we trained the models on Arabic dialects before retraining them specifically on the Algerian dialect. The results we obtained were quite promising, as the bleu score value of the seq2seq model increased by +34 points, while that of the attentional seq2seq model increased by +19 points.

Although the attention mechanism did not have a significant impact on the performance of transfer learning in experiment 2-TL, it was still an essential component of the model. This was made evident by the fact that the attention model's final results were better than those of the model without attention.

Additionally, in experiment 5-ITL, 6-ITL, and 7-ITL, we observed that the transformer model, incorporating attention mechanisms in both its encoder and decoder, surpassed the performance of the attentional seq2seq model. By incorporating attention mechanisms at

multiple levels, the transformer model demonstrated a remarkable ability to comprehend and generate coherent sequences, leading to notably improved outcomes.

All in all, our experiment provided valuable insight into the effectiveness of attention mechanisms in NMT models, particularly in the context of transfer learning.

The Impact of Sentence Length on Transfer Learning NMT for Arabic Dialects

Figure 5.9 displays the BLEU scores for seven different experiments, each varying in sentence size. Upon examining Experiment 2-TL, it is noted that the seq2seq and attentional seq2seq models perform relatively well for sentence sizes between 10 and 25, with scores ranging from 56.87 to 73.30. However, for shorter sentences (0-5), the models perform poorly, yielding a score of only 11.21.

Experiment 5-TL demonstrates better performance for small sentences, with scores ranging from 1.20 to 2.24 for sentence sizes between 5 and 20, but lower scores for longer sentences (20-25), for both model attentional seq2seq and transformer.

Experiment 6-TL presents comparable results to Experiment 5-TL, with low scores for very short and long sentences, and higher scores for sentence sizes between 5 and 15, for both models.

Experiment 7-TL displays low scores for short sentences with attentional seq2seq, but performs better for sentence sizes between 10 and 20, with both models. However, the models perform notably poorly for sentences of length 20-25.

Experiment 5-ITL yields relatively favorable scores for sentence sizes between 0 and 25 with attentional seq2seq, and between 0 and 15 with transformer, with scores ranging from 1.77 to 11.22.

Experiment 6-ITL yields results that are similar to Experiment 5-ITL, exhibiting higher scores for sentence sizes ranging from 5 to 25, using both models. However, for very short sentences, the scores are lower in the case of the attentional seq2seq model.

Lastly, Experiment 7-ITL showcases consistent scores across all sentence sizes when employing the attentional seq2seq model. Notably, the model exhibits exceptional performance for lengthier sentences, achieving impressive scores ranging from 2.94 to 9.10 for sentence sizes between 15 and greater than 25 words. On the other hand, the transformer model delivers results akin to those observed in Experiments 5-ITL and 6-ITL, implying a comparable outcome in this context.

In summary, the NMT models using transfer learning strategy exhibit better performance for sentence sizes ranging from 5 to 25, with scores ranging from 1.77 to 73.30. These sentence sizes represent a moderate range, not too short or too long.

In conclusion, this study highlights that sentence length can impact NMT model performance. Specifically, the sweet spot for these NMT models is for sentence sizes between 5 and 25. It's important to note that these results are based on the specific experiments and



Figure 5.9: BLEU score value relative to the length of the for each Transfer learning experiment.

models presented in the table and that the performance of NMT models can differ based on various factors, such as the language pair, domain, and training data.

5.5 Conclusion

This chapter has provided an overview of transfer learning and highlighted its potential for improving the accuracy of MT systems. We discussed transfer learning, setting, different approaches, and how it works, particularly in the context of improving NMT performance for low-resource languages such as Arabic dialects.

Furthermore, we presented our contribution to the field of Arabic dialect NMT, where we applied transfer learning to improve the accuracy of Maghrebi NMT while using a smaller training set. This approach provides a potential solution to the problem of the limited availability of data for low-resource languages, such as Arabic dialects, and presents a valuable addition to the field of NMT.

Overall, the approaches and methods discussed in this chapter demonstrate the potential of transfer learning in NMT and its ability to significantly impact the development of more accurate and effective machine translation systems. We believe that our contribution and proposed approach can serve as a foundation for further research in this area and pave the way for the development of more advanced and robust MT systems.

6

Conclusion

The objective of this thesis is to explore the application of artificial intelligence approaches to the translation of Algerian dialect. With the increasing globalization and mobility, there is a growing need for effective communication between people from different linguistic backgrounds. However, the translation of dialects, especially those with limited resources and linguistic features, remains a challenging task. This thesis seeks to address this challenge by investigating the use of artificial intelligence, particularly NMT, to improve the translation quality of the Algerian dialect. By developing and testing various models and techniques, this thesis aims to contribute to the advancement of the field of MT and a better understanding of the linguistic features and characteristics of the Algerian dialect.

Summary of Main Findings and Contributions

Our research aimed to address the significant challenge of translating Algerian dialects using NMT models. To achieve this ambitious objective, we extensively utilized models such as seq2seq, Attentional seq2seq, and Transformer models. Furthermore, we relied upon two multi-dialectal Arabic parallel corpora, namely PADIC and MADAR, which served as our primary data sources. Moreover, we conducted thorough investigations into the profound impact of the attention mechanism on the translation of diverse Arabic dialects, including Algerian, Tunisian, and Moroccan dialects. By meticulously designing multiple experiments, each with its own unique set of experimental conditions, we were able to gain deep insights into the intricacies of dialect translation. Our evaluations were

rigorously carried out using the widely used BLEU score metric, which served as a reliable yardstick to gauge the performance and effectiveness of our system.

Our experiments yielded the following results:

- seq2seq model, employing data augmentation, yielded a BLEU score of 22.10 points for the Algerian dialect based on PADI
- 0.05 BLEU score points for Algerian dialect from PADIC using the seq2seq model
- 16.5 BLEU score points for Algerian dialect from PADIC using the attentional seq2seq model
- 33.58 BLEU score points for Arabic dialect from PADIC using the seq2seq model.
- 54.49 BLEU score points for Arabic dialect from MADAR using the seq2seq model.
- 44.53 BLEU score points for Arabic dialect from PADIC using the attentional seq2seq model.
- 56.10 BLEU score points for Arabic dialect from MADAR using the attentional seq2seq model.
- 0.05 BLEU score points for Algerian dialect from PADIC and MADAR using the attentional seq2seq model
- 0.03 BLEU score points for Tunisian dialect from PADIC and MADAR using the attentional seq2seq model
- 0.04 BLEU score points for Moroccan dialect from PADIC and MADAR using the attentional seq2seq model
- 1.23 BLEU score points for Algerian dialect from PADIC and MADAR using the Transformer model
- 1.26 BLEU score points for Tunisian dialect from PADIC and MADAR using the Transformer model
- 1.01 BLEU score points for Moroccan dialect from PADIC and MADAR using the Transformer model

The results of our experiments provide compelling evidence of the significant positive impact exerted by the attention mechanism on the performance of our system. In fact, we observed an impressive average improvement of approximately 7.4 points in the BLEU score for Arabic dialects originating from both the PADIC and the MADAR corpora. This

substantial enhancement clearly underscores the crucial role played by the attention mechanism in enhancing the translation quality of Algerian dialect.

Moreover, our investigation into the Transformer model, which extensively relies on the attention mechanism in both its encoder and decoder components, yielded even more remarkable outcomes. When compared to the attentional seq2seq model, we noticed a consistent and noteworthy increase in the BLEU score for each of the Algerian, Tunisian, and Moroccan dialects, utilizing data from both the PADIC and the MADAR corpora. This compelling finding serves as strong evidence of the superior capabilities of the attention mechanism in effectively capturing the intricate linguistic nuances inherent in these dialects and subsequently generating more accurate and contextually appropriate translations.

While the attention mechanism exhibits commendable positive effects on our translation system, it is crucial to recognize that these advantages do not overshadow the substantial negative influence of poverty within the data. This challenge is further compounded by the presence of rare words and the linguistic complexities inherent in Arabic dialects, creating a formidable hurdle for translation systems to overcome.

The scarcity of training data poses a significant challenge when dealing with the linguistic features of Arabic dialects, which encompass a wide range of vocabulary, idiomatic expressions, and region-specific grammatical structures. Limited data availability hinders translation systems from accurately capturing and understanding these nuanced elements. Moreover, the presence of rare words in small datasets exacerbates the problem by making it difficult for models to learn their correct translations and incorporate them effectively. This limitation impairs the system's ability to generate accurate and contextually appropriate translations, particularly in the complex linguistic landscape of Arabic dialects.

In our efforts to enhance the performance of our specific low-resource dialects translation model, such as Algerian, Tunisian, and Moroccan dialects, we aimed to explore alternative methods beyond simply increasing the amount of data. One approach we adopted was transfer learning, which involves utilizing existing knowledge from related tasks to improve performance on a specific task.

To implement this approach, we sought to leverage the shared linguistic features between the specific low-resource and Arabic dialects. By transferring knowledge from the more widely-used Arabic dialects to the less-represented specific low-resource dialects, we aimed to enhance the model's understanding of the commonalities between Arabic and specific low-resource dialects, and enable the model to specialize in the unique features of specific low-resource dialects.

We opted to use the transductive transfer learning scenario, which entails identifying specific instances within the dataset that are similar to the target dialects. We used the instance-based transfer learning technique to identify these instances, which allowed us to build a more specialized model that was better able to capture the nuances of the specific low-resource dialects. Overall, our use of transfer learning provided an effective means of

improving the performance of our translation model beyond merely increasing the amount of available data.

For conducting our experiments, we employed seq2seq, attentional seq2seq, and transformer models. As a baseline experiment, we utilized a previous learning experiment to assess the effectiveness of our strategy. The results were remarkable, with the seq2seq model yielding a significant increase of +34 points in the BLEU score for the Algerian dialect, while the attentional seq2seq model achieved a notable increase of +19 points.

In additional experiments focusing on Algerian, Tunisian, and Moroccan dialects, we observed a further improvement in translation quality. The attentional seq2seq and transformer models exhibited an increase of +1 point in the BLEU score for these dialects.

Following our initial success with transfer learning from Arabic dialects to low resource Algerian, Tunisian, and Moroccan dialects, we continued to explore the potential of this approach. We sought to further improve the performance of our model by transferring knowledge through multiple levels, beginning with Arabic dialects and progressing to Maghreb dialects and eventually to specific low-resource dialects, such as Algerian, Tunisian, and Moroccan dialects.

To achieve this, we proposed an incremental transfer learning strategy, in which knowledge is transferred incrementally between increasingly specific dialects. This allowed us to build on the knowledge gained from transferring between broader dialect categories and apply it to the unique features of individual dialects. Our experiments demonstrated that our approach, which utilized the attentional seq2seq and transformer models, contributed to an increase in BLEU scores. In particular, we observed a +2 points with attentional seq2seq and +5 points with transformer increase in the BLEU score for the Algerian dialect and a +1 point increase for both Tunisian and Moroccan dialects, with both models.

Overall, our incremental transfer learning approach proved to be a powerful strategy for improving the performance of our model on low-resource dialects, such as those found in the Maghreb region. Through this approach, we were able to gradually build a more specialized and accurate model, leading to significant improvements in translation quality.

Insights on a practical application

The thesis at hand aims to address the challenges faced in translating Arabic dialects into MSA using NMT models. Arabic dialects pose unique challenges for MT due to their limited resources and standardization. To overcome these challenges, the study utilizes both the seq2seq, attentional seq2seq, and transformer models, and compares the results to determine the impact of attention on the translation quality of Arabic dialects. By doing so, the research provides valuable insights into the effectiveness of different NMT models for Arabic dialects and contributes to the development of Arabic language technology.

Furthermore, the study highlights the potential of NMT models in improving the translation of Arabic dialects into MSA. It emphasizes the need for continued research in this field, as these dialects suffer from a lack of appropriate data for machine learning models. The study also explores the use of transfer learning, which involves transferring common knowledge among all Arabic dialects to a special dialect with limited resources. This method can help to improve the accuracy of the model by transferring the shared features in various Arabic dialects to a specific dialect with limited resources.

To enhance the accuracy of the MT models in translating Arabic dialects into MSA, the study went further by exploring the optimal sentence lengths for learning. Through this investigation, the study aimed to identify the most appropriate sentence lengths for learning that would be beneficial for the translation model.

Overall, the thesis presents a comprehensive analysis of the challenges faced in translating Arabic dialects into MSA and offers potential solutions using NMT models. Its findings have important implications for the promotion of communication and cultural exchange in the Arab world and beyond. By developing more accurate and effective MT models, the study contributes to bridging the linguistic divide between different Arabic dialects and promoting mutual understanding and cooperation.

Limitations and Future Directions

The scarcity of suitable data poses a significant hurdle for NMT, particularly in the context of Arabic dialects. The limited availability of appropriate training data impedes NMT models' ability to precisely translate these dialects into MSA or other languages. This challenge is particularly pronounced in the case of the Algerian dialect and other Arabic dialects, given their unique characteristics and the difficulty in obtaining sufficient data for model training. Consequently, it is imperative to devise effective strategies to overcome the constraints posed by the lack of data for Arabic dialects, ensuring the accuracy and efficacy of NMT models.

Our research findings revealed modest outcomes, attributable to the utilization of small databases that inadequately capture the diverse linguistic features inherent in Arabic dialects. Recognizing the utmost importance of creating comprehensive and extensive databases that encompass the full spectrum of linguistic characteristics within these dialects is pivotal.

In considering the trajectory of our work, two promising avenues emerge. One entails leveraging generative artificial intelligence to construct an extensive corpus of Arabic dialects, thereby significantly enriching the linguistic resources specific to these dialects. Conversely, a focus on a transfer learning strategy involves harnessing pre-trained large language models, enabling the efficient transfer of knowledge and its customization to the unique requirements of the Arabic dialect context. This approach has the potential to

streamline development processes and enhance overall performance.

Addressing the challenge of insufficient data for Arabic dialects is paramount for advancing neural machine translation, fostering communication, and promoting cultural exchange in the Arab world and beyond. This necessitates ongoing research efforts and the development of innovative strategies capable of overcoming the limitations imposed by the lack of data, ultimately enabling more effective and accurate translations of Arabic dialects.

References

- [1] Janet Watson. Arabic dialects (general article). 2011.
- [2] Salima Harrat, Karima Meftouh, and Kamel Smaili. Machine translation for arabic dialects (survey). *Information Processing & Management*, 56(2):262–273, 2019.
- [3] Karima Meftouh, Salima Harrat, Salma Jamoussi, Mourad Abbas, and Kamel Smaili. Machine translation experiments on padic: A parallel arabic dialect corpus. In *Proceedings of the 29th Pacific Asia conference on language, information and computation*, pages 26–34, 2015.
- [4] Fatiha Sadat, Fatma Mallek, Mohamed Mahdi Boudabous, Rahma Sellami, and Atefeh Farzindar. Collaboratively constructed linguistic resources for language variants and their exploitation in nlp application—the case of tunisian arabic and the social media. In *Proceedings of workshop on Lexical and grammatical resources for language processing*, pages 102–110, 2014.
- [5] Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar Zaidan, and Chris Callison-Burch. Machine translation of arabic dialects. In *Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pages 49–59, 2012.
- [6] W Aransa. *Statistical Machine Translation of the Arabic Dialect*. PhD thesis, University of Maine, doctoral school STIM, 2015.
- [7] Hitham Abo Bakr, Khaled Shaalan, and Ibrahim Ziedan. A hybrid approach for converting written egyptian colloquial dialect into diacritized arabic. In *The 6th international conference on informatics and systems, infos2008, cairo university*, 2008.
- [8] Jonathan May, Yassine Benjira, and Abdessamad Echihabi. An arabizi-english social media statistical machine translation system. In *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: MT Researchers Track*, pages 329–341, 2014.

- [9] Nizar Y Habash. *Introduction to Arabic natural language processing*. Synthesis Lectures on Human Language Technologies (SLHLT), Springer Nature, New York, 2010.
- [10] Omar F Zaidan and Chris Callison-Burch. Arabic dialect identification. *Computational Linguistics*, 40(1):171–202, 2014.
- [11] Faical Azouaou and Imane Guellil. Alg/fr: A step by step construction of a lexicon between algerian dialect and french. In *The 31st Pacific Asia Conference on Language, Information and Computation PACLIC*, volume 31, 2017.
- [12] CHAMI Abdelkarim. A historical background of the linguistic situation in algeria. *Almawaqif Journal*, 1:387–395, 2009.
- [13] Khaoula Taleb Ibrahimi. L’algérie: coexistence et concurrence des langues. *L’année du Maghreb*, (I):207–218, 2006.
- [14] History of algeria. <https://www.embalgeria.nl/history-of-algeria/>. Accessed: May 18, 2023.
- [15] Abd El-Aziz Mohammed Hassan. علم اللغة الاجتماعي. Egyptian Publishing House, 2009.
- [16] Nouiouat Mokhtar El-ahmedi. عن اللسان و في البيان: مقالات و افتتاحيات. Supreme Council for the Arabic Language Publishing House, 2020.
- [17] TASSIST Sid Ahmed. Multiplicité des confluent linguistiques dans le parler algérien multiplicity of linguistic confluences in the algerian spoken language.
- [18] Fatema Kajaot. الاختلافات اللهجية في المنطوق الجزائرية: دراسة صوتية مقارنة. University of Ahmed Zabana Relizane, 2020.
- [19] Noureddine Guella. Emprunts lexicaux dans des dialectes arabes algériens. *Synergies Monde arabe*, 8:81–88, 2011.
- [20] El-Abodi Mohammed Ben-Nacer. الكلمات الدخيلة في لغتنا الدارجة. King Abdulaziz Public Library, 2005.
- [21] Faraa Chahda, Hamdane Jihad, Amamria Moussa, and El-Annabi Mohammed. مقدمة في اللغويات المعاصرة. Wael Publishing House, 2015.
- [22] Hamida Aouaidjia. *Assessment of the oral language of a deaf child with a cochlear implant: a proposal to construct a test to assess auditory perception and verbal expression*. PhD thesis, University of Badji Mokhtar Annaba.

- [23] Ahmed Said, Mohamed El-Sharqwi, Achraf Chalabi, and Eslam Kamal. A hybrid approach for arabic diacritization. In *International Conference on Application of Natural Language to Information Systems*, pages 53–64. Springer, 2013.
- [24] Salima Harrat, Karima Meftouh, Mourad Abbas, Walid-Khaled Hidouci, and Kamel Smaili. An algerian dialect: Study and resources. *International journal of advanced computer science and applications (IJACSA)*, 7(3):384–396, 2016.
- [25] Mohammed Nasser. Processes of word formation in english and arabic. *Journal of the College of Education, Babylon University, Iraq*, 2(3):71–87, 2008.
- [26] A .Z Obiedat. Axiomatizing elementary arabic syntax. *International Journal for Arabic Linguistics and Literature Studies*, 1(1):73–89, 2019.
- [27] MD Okpor. Machine translation approaches: issues and challenges. *International Journal of Computer Science Issues (IJCSI)*, 11(5):159, 2014.
- [28] Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. A study of reinforcement learning for neural machine translation. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun’ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3612–3621, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.
- [29] John Hutchins. The history of machine translation in a nutshell. *Retrieved December*, 20(2009):1–1, 2005.
- [30] P J Antony. Machine translation approaches and survey for indian languages. *International journal of computational Linguistics and Chinese Language Processing*, 18(1):47–78, 2013.
- [31] Benson Kituku, Lawrence Muchemi, and Wanjiku Nganga. A review on machine translation approaches. *Indonesian Journal of Electrical Engineering and Computer Science*, 1(1):182–190, 2016.
- [32] Jože Kadivec. *Prilagoditev statističnega strojnega prevajalnika za specifično domeno v slovenskem jeziku*. PhD thesis, Univerza v Ljubljani, 2016.
- [33] Peng Li. A survey of machine translation methods. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 11(12):7125–7130, 2013.
- [34] Harold Somers. An introduction to machine translation. 1992.
- [35] Jonathan Slocum. A survey of machine translation: Its history, current status and future prospects. *Computational linguistics*, 11(1):1–17, 1985.

- [36] Thierry Poibeau. *Machine Translation*. The MIT Press Essential Knowledge, 2017.
- [37] Ralph Grishman. *Computational linguistics: an introduction*. Cambridge University Press, 1986.
- [38] Richard H Richens. Interlingual machine translation. *The Computer Journal*, 1(3):144–147, 1958.
- [39] Zhou Dajun and Wang Yun. Corpus-based machine translation: Its current development and perspectives. In *International Forum of Teaching and Studies*, volume 11, page 90. American Scholars Press, Inc., 2015.
- [40] Adam Lopez. Statistical machine translation. *ACM Computing Surveys (CSUR)*, 40(3):1–49, 2008.
- [41] Makoto Nagao. *A framework of a mechanical translation between Japanese and English by analogy principle*. Elsevier Science Publishers, 1984.
- [42] Harold Somers. Example-based machine translation. *Machine translation*, 14:113–157, 1999.
- [43] John Hutchins. Example-based machine translation: a review and commentary. *Machine Translation*, 19(3):197–211, 2005.
- [44] Sandeep Saini and Vineet Sahula. A survey of machine translation techniques and systems for indian languages. In *2015 IEEE International Conference on Computational Intelligence & Communication Technology*, pages 676–681. IEEE, 2015.
- [45] Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1700–1709, 2013.
- [46] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [47] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.
- [48] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.

- [49] Zhixing Tan, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun, and Yang Liu. Neural machine translation: A review of methods, resources, and tools. *AI Open*, 1:5–21, 2020.
- [50] Marcin Junczys-Dowmunt, Tomasz Dwojak, and Hieu Hoang. Is neural machine translation ready for deployment? a case study on 30 translation directions. In Mauro Cettolo, Jan Niehues, Sebastian Stüker, Luisa Bentivogli, Rolando Cattoni, and Marcello Federico, editors, *Proceedings of the 13th International Conference on Spoken Language Translation*, Seattle, Washington D.C, December 8-9 2016. International Workshop on Spoken Language Translation.
- [51] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.
- [52] Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, et al. Achieving human parity on automatic chinese to english news translation. *arXiv preprint arXiv:1803.05567*, 2018.
- [53] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. Understanding the exploding gradient problem. *CoRR*, abs/1211.5063, 2(417):1, 2012.
- [54] Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, Jürgen Schmidhuber, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. 2001.
- [55] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [56] Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder–decoder approaches. In Dekai Wu, Marine Carpuat, Xavier Carreras, and Eva Maria Vecchi, editors, *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [57] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- [58] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In Lluís Màrquez, Chris Callison-Burch,

- and Jian Su, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal, September 2015. Association for Computational Linguistics.
- [59] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *International conference on machine learning*, pages 1243–1252. PMLR, 2017.
- [60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [61] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [62] Ankur Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. In Jian Su, Kevin Duh, and Xavier Carreras, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2249–2255, Austin, Texas, November 2016. Association for Computational Linguistics.
- [63] Mia Xu Chen, Orhan Firat, Ankur Bapna, Melvin Johnson, Wolfgang Macherey, George Foster, Llion Jones, Mike Schuster, Noam Shazeer, Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Lukasz Kaiser, Zhifeng Chen, Yonghui Wu, and Macduff Hughes. The best of both worlds: Combining recent advances in neural machine translation. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 76–86, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [64] Chris Callison-Burch, Cameron Shaw Fordyce, Philipp Koehn, Christof Monz, and Josh Schroeder. (meta-) evaluation of machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 136–158, 2007.
- [65] John S White and Theresa A O’Connell. Evaluation in the arpa machine translation program: 1993 methodology. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey*, 1994.
- [66] Franz Josef Och. Minimum error rate training in statistical machine translation. In *Proceedings of the 41st annual meeting of the Association for Computational Linguistics*, pages 160–167, 2003.

- [67] 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*, pages 311–318, 2002.
- [68] Bonnie Dorr, Matt Snover, and Nitin Madnani. Part 5: machine translation evaluation. *Handb. Nat. Lang. Process. Mach. Transl. DARPA Glob. Auton. Lang. Exploit*, page 936, 2011.
- [69] Satanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72, 2005.
- [70] Christiane Fellbaum. A semantic network of english: the mother of all wordnets. In *EuroWordNet: A multilingual database with lexical semantic networks*, pages 137–148. Springer, 1998.
- [71] Houda Bouamor, Nizar Habash, Mohammad Salameh, Wajdi Zaghouani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann, et al. The madar arabic dialect corpus and lexicon. In *LREC*, 2018.
- [72] Houda Bouamor, Nizar Habash, and Kemal Oflazer. A multidialectal parallel corpus of arabic. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, pages 1240–1245, 2014.
- [73] Hamdy Mubarak. Dial2msa: a tweets corpus for converting dialectal arabic to modern standard arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), OSACT2018 Workshop*, pages 49–53, 2018.
- [74] Ibrahim Gashaw and HL Shashirekha. Amharic-arabic neural machine translation. *arXiv preprint arXiv:1912.13161*, 2019.
- [75] Amjad Almahairi, Kyunghyun Cho, Nizar Habash, and Aaron Courville. First result on arabic neural machine translation. *arXiv preprint arXiv:1606.02680*, 2016.
- [76] Ebtesam H Almansor and Ahmed Al-Ani. A hybrid neural machine translation technique for translating low resource languages. In *Machine Learning and Data Mining in Pattern Recognition: 14th International Conference, MLDM 2018, New York, NY, USA, July 15-19, 2018, Proceedings, Part II 14*, pages 347–356. Springer, 2018.

- [77] Imane Guellil, Faïçal Azouaou, Mourad Abbas, and Sadat Fatiha. Arabizi transliteration of algerian arabic dialect into modern standard arabic. *Social MT*, pages 1–8, 2017.
- [78] El Moatez Billah Nagoudi, AbdelRahim Elmadany, and Muhammad Abdul-Mageed. Investigating code-mixed modern standard arabic-egyptian to english machine translation. *arXiv preprint arXiv:2105.13573*, 2021.
- [79] Amel Slim, Ahlem Melouah, Yousef Faghihi, and Khouloud Sahib. Algerian dialect translation applied on covid-19 social media comments. In *Artificial Intelligence and Renewables Towards an Energy Transition 4*, pages 716–726. Springer, 2021.
- [80] Amel Slim, Ahlem Melouah, Usef Faghihi, and Khouloud Sahib. Improving neural machine translation for low resource algerian dialect by transductive transfer learning strategy. *Arabian Journal for Science and Engineering*, 47(8):10411–10418, 2022.
- [81] Amel Slim, Ahlem Melouah, Yousef Faghihi, and Khouloud Sahib. The impact of attention mechanism on arabic dialect neural machine translation. In *ICCSAITCS ‘2022: International Conference on Cyber Security, Artificial Intelligence and Theoretical Computer ScienceAt: Boumrdes, Algeria*, 2022.
- [82] Jonas Waldendorf, Alexandra Birch, Barry Hadow, and Antonio Valerio Micele Barone. Improving translation of out of vocabulary words using bilingual lexicon induction in low-resource machine translation. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 144–156, 2022.
- [83] Junjie Hu, Mengzhou Xia, Graham Neubig, and Jaime Carbonell. Domain adaptation of neural machine translation by lexicon induction. *arXiv preprint arXiv:1906.00376*, 2019.
- [84] Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. Neural machine translation for low-resource languages: A survey. *CoRR*, abs/2106.15115, 2021.
- [85] Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5):1–38, 2020.
- [86] Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O.K. Li. Universal neural machine translation for extremely low resource languages. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

- Technologies, Volume 1 (Long Papers)*, pages 344–354, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [87] Minh-Thang Luong, Quoc V. Le, Ilya Sutskever, Oriol Vinyals, and Lukasz Kaiser. Multi-task sequence to sequence learning. *CoRR*, abs/1511.06114, 2015.
- [88] Barret Zoph and Kevin Knight. Multi-source neural translation. In Kevin Knight, Ani Nenkova, and Owen Rambow, editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 30–34, San Diego, California, June 2016. Association for Computational Linguistics.
- [89] Thanh-Le Ha, Jan Niehues, and Alex Waibel. Toward multilingual neural machine translation with universal encoder and decoder. In Mauro Cettolo, Jan Niehues, Sebastian Stüker, Luisa Bentivogli, Rolando Cattoni, and Marcello Federico, editors, *Proceedings of the 13th International Conference on Spoken Language Translation*, Seattle, Washington D.C, December 8-9 2016. International Workshop on Spoken Language Translation.
- [90] Graeme Blackwood, Miguel Ballesteros, and Todd Ward. Multilingual neural machine translation with task-specific attention. *arXiv preprint arXiv:1806.03280*, 2018.
- [91] Yining Wang, Jiajun Zhang, Feifei Zhai, Jingfang Xu, and Chengqing Zong. Three strategies to improve one-to-many multilingual translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2955–2960, 2018.
- [92] Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. Google’s multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351, 2017.
- [93] Toan Q. Nguyen and David Chiang. Transfer learning across low-resource, related languages for neural machine translation. In Greg Kondrak and Taro Watanabe, editors, *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 296–301, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.
- [94] Raj Dabre, Tetsuji Nakagawa, and Hideto Kazawa. An empirical study of language relatedness for transfer learning in neural machine translation. In *Proceedings of*

- the 31st Pacific Asia conference on language, information and computation*, pages 282–286, 2017.
- [95] Tom Kocmi and Ondřej Bojar. Trivial transfer learning for low-resource neural machine translation. In Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Christof Monz, Matteo Negri, Aurélie Névél, Mariana Neves, Matt Post, Lucia Specia, Marco Turchi, and Karin Verspoor, editors, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 244–252, Brussels, Belgium, October 2018. Association for Computational Linguistics.
- [96] Yong Cheng. Joint training for pivot-based neural machine translation. In *Joint training for neural machine translation*, pages 41–54. Springer, 2019.
- [97] Shuo Ren, Wenhua Chen, Shujie Liu, Mu Li, Ming Zhou, and Shuai Ma. Triangular architecture for rare language translation. In Iryna Gurevych and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 56–65, Melbourne, Australia, July 2018. Association for Computational Linguistics.
- [98] Chen Zhang, Xu Tan, Yi Ren, Tao Qin, Kejun Zhang, and Tie-Yan Liu. Uwspeech: Speech to speech translation for unwritten languages. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14319–14327, 2021.
- [99] Anna Currey and Kenneth Heafield. Zero-resource neural machine translation with monolingual pivot data. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 99–107, 2019.
- [100] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [101] Andrew Arnold, Ramesh Nallapati, and William W Cohen. A comparative study of methods for transductive transfer learning. In *Seventh IEEE international conference on data mining workshops (ICDMW 2007)*, pages 77–82. IEEE, 2007.
- [102] Q Y Zhang, W Dai, and S J Pan. *Transfer learning*. Cambridge University Press, 2020.
- [103] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big data*, 3(1):1–40, 2016.
- [104] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

- [105] Yi Yao and Gianfranco Doretto. Boosting for transfer learning with multiple sources. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1855–1862. IEEE, 2010.
- [106] Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu. Eigentransfer: a unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 193–200, 2009.
- [107] Jing Jiang and ChengXiang Zhai. Instance weighting for domain adaptation in nlp. *ACL*, 2007.
- [108] Masashi Sugiyama, Shinichi Nakajima, Hisashi Kashima, Paul Buenau, and Motoaki Kawanabe. Direct importance estimation with model selection and its application to covariate shift adaptation. *Advances in neural information processing systems*, 20, 2007.
- [109] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng. Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on Machine learning*, pages 759–766, 2007.
- [110] Hal Daumé III. Frustratingly easy domain adaptation. In Annie Zaenen and Antal van den Bosch, editors, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [111] Sinno Jialin Pan, James T Kwok, Qiang Yang, et al. Transfer learning via dimensionality reduction. In *AAAI*, volume 8, pages 677–682, 2008.
- [112] Wenyuan Dai, Qiang Yang, Gui-Rong Xue, and Yong Yu. Self-taught clustering. In *Proceedings of the 25th international conference on Machine learning*, pages 200–207, 2008.
- [113] Zheng Wang, Yangqiu Song, and Changshui Zhang. Transferred dimensionality reduction. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part II 19*, pages 550–565. Springer, 2008.
- [114] Li Fei-Fei, Robert Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006.
- [115] Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In

- 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3081–3088. IEEE, 2010.
- [116] Luo Jie, Tatiana Tommasi, and Barbara Caputo. Multiclass transfer learning from unconstrained priors. In *2011 International Conference on Computer Vision*, pages 1863–1870. IEEE, 2011.
- [117] Jesse Davis and Pedro Domingos. Deep transfer via second-order markov logic. In *Proceedings of the 26th annual international conference on machine learning*, pages 217–224, 2009.
- [118] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62:107–136, 2006.
- [119] Laith H Baniata, Seyoung Park, Seong-Bae Park, et al. A neural machine translation model for arabic dialects that utilizes multitask learning (mtl). *Computational intelligence and neuroscience*, 2018, 2018.
- [120] Youness Moukafih, Nada Sbihi, Mounir Ghogho, and Kamel Smaïli. Improving machine translation of arabic dialects through multi-task learning. In *AIXIA 2021–Advances in Artificial Intelligence: 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, December 1–3, 2021, Revised Selected Papers*, pages 580–590. Springer, 2022.
- [121] Laith H Baniata, Seyoung Park, and Seong-Bae Park. A multitask-based neural machine translation model with part-of-speech tags integration for arabic dialects. *Applied Sciences*, 8(12):2502, 2018.
- [122] Wael Farhan, Bashar Talafha, Analle Abuammar, Ruba Jaikat, Mahmoud Al-Ayyoub, Ahmad Bisher Tarakji, and Anas Toma. Unsupervised dialectal neural machine translation. *Information Processing & Management*, 57(3):102181, 2020.
- [123] Yunsu Kim, Yingbo Gao, and Hermann Ney. Effective cross-lingual transfer of neural machine translation models without shared vocabularies. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1246–1257, Florence, Italy, July 2019. Association for Computational Linguistics.
- [124] Shuteng Niu, Yongxin Liu, Jian Wang, and Houbing Song. A decade survey of transfer learning (2010–2020). *IEEE Transactions on Artificial Intelligence*, 1(2):151–166, 2020.
- [125] Clive Holes. *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press, 2004.

- [126] Houda Saadane and Nizar Habash. A conventional orthography for algerian arabic. In *the Second Workshop on Arabic Natural Language Processing*, pages 69–79, 2015.
- [127] Imane Guellil, Houda Saâdane, Faical Azouaou, Billel Gueni, and Damien Nouvel. Arabic natural language processing: An overview. *Journal of King Saud University-Computer and Information Sciences*, 33(5):497–507, 2021.
- [128] Nizar Habash, Mona T Diab, and Owen Rambow. Conventional orthography for dialectal arabic. In *LREC*, pages 711–718, 2012.
- [129] Inès Zribi, Mariem Ellouze, Lamia Hadrich Belguith, and Philippe Blache. Morphological disambiguation of tunisian dialect. *Journal of king Saud University-computer and information sciences*, 29(2):147–155, 2017.
- [130] Salima Harrat, Karima Meftouh, and Kamel Smaïli. Maghrebi arabic dialect processing: an overview. *Journal of International Science and General Applications*, 1, 2018.
- [131] Abdessatar Mahfoudhi. Agreement lost, agreement regained: A minimalist account of word order and agreement variation in arabic. *California Linguistic Notes*, 27(2):1–28, 2002.
- [132] Mostafa Lameen Souag. Explorations in the syntactic cartography of algerian arabic. *Unpublished MA*), *University of London, School of Oriental and African Studies, London*, 2006.
- [133] Moha Ennaji. *Multilingualism, cultural identity, and education in Morocco*. Springer Science & Business Media, 2005.

List of Publications and Presentations

Journals Paper

1. A. Slim, A. Melouah, U. Faghihi, and K. Sahib. Improving Neural Machine Translation for Low Resource Algerian Dialect by Transductive Transfer Learning Strategy, *Arab J Sci Eng*, 2022, **47(8)**, 10411–10418.

BookChapter

1. A. Slim, A. Melouah, U. Faghihi, and K. Sahib. Algerian Dialect Translation Applied on COVID-19 Social Media Comments. *Artificial Intelligence and Renewables Towards an Energy Transition.*, In: Hatti, M. (eds), ICAIRES 2020. Lecture Notes in Networks and Systems, vol 174. Springer, Cham, (2021).

Conference Paper

1. A. Slim, A. Melouah, U. Faghihi, and K. Sahib, The Impact of Attention Mechanism on Arabic Dialect Neural Machine Translation, *ICCSAITCS '2022: International Conference on Cyber Security, Artificial Intelligence and Theoretical Computer Science*, Boumrdes, Algeria, December 2022.