BIDIRECTIONAL LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK BASED MYANMAR DIALOGUE ACT RECOGNITION

SANN SU SU YEE

UNIVERSITY OF COMPUTER STUDIES, YANGON

JUNE, 2024

Bidirectional Long Short-Term Memory Recurrent Neural Network Based Myanmar Dialogue Act Recognition

Sann Su Su Yee

University of Computer Studies, Yangon

A thesis submitted to the University of Computer Studies, Yangon in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

June, 2024

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

…..…… ……..……

Date Sann Su Su Yee

ACKNOWLEDGEMENTS

First and foremost, I would like to thank His Excellency, the Minister, the Ministry of Science and Technology for full facilities support during the Ph. D Course at the University of Computer Studies, Yangon.

I would like to express very special gratitude goes to Dr. Mie Mie Khin, the Rector, the University of Computer Studies, Yangon, and Dr. Mie Mie Thet Thwin, the former Rector, the University of Computer Studies, Yangon, for allowing me to develop this thesis and giving me general guidance during the period of my study.

I would also like to extend my special appreciation and thanks to Dr. Si Si Mar Win, Professor, Dr. Tin Thein Thwel and Dr. Sapal Phyu, former Professor, and Course-coordinator of the Ph.D 11th Batch, the University of Computer Studies, Yangon, for their useful comments, advice and insight which are invaluable to me.

I sincerely would like to express my deepest gratitude to my supervisor, Dr. Khin Mar Soe, Professor, the University of Computer Studies, Yangon. Without her excellent guidance, caring, patience, and persistent help, this dissertation would not have been possible.

I deeply would like to express my respectful gratitude to Dr. Win Pa Pa, Professor, the University of Computer Studies, Yangon, for her valuable advice, many insightful discussion and suggestions in my research work.

I am indeed obliged to Daw Aye Aye Khine, Professor, Head of English Department for her valuable supports from the language point of view and pointed out the correct usage in my dissertation.

I also thank my classmates from Ph.D 11th batch and my colleagues in Natural Language Processing Lab., for providing support and friendship that I needed.

I am very much indebted to my parents, my brothers, and my sister-in-law for always believing in me, for their endless love and support. They are always supporting and encouraging me during the years of my Ph.D study.

ABSTRACT

This research aims to develop the deep learning-based Myanmar Dialogue Act Recognition (MDAR) system to enhance Myanmar Dialogue Systems. Dialogue Act (DA) recognition is a foundational aspect of dialogue understanding, capturing user intent at the sentence level with units such as greeting, question, and inform. By identifying these intents, dialogue systems can interact more naturally and effectively with users. This study explores current approaches to DA recognition, specifically focusing on Myanmar dialogues, a previously underrepresented area in Natural Language Processing (NLP) research. Initially, two machine learning techniques—Naïve Bayes classifier and Support Vector Machine (SVM)—were applied to the MmTravel corpus, a dataset comprising Myanmar travel-related conversations. Both approaches demonstrated moderately good accuracy for Myanmar dialogue tagging, with SVM showing a slightly better performance.

Recognizing the critical role of Spoken Language Understanding (SLU) in dialogue systems, this research emphasizes DA recognition as an essential pre-processing step for speech understanding. To further improve DA recognition accuracy, this research proposes a deep learning-based DA model utilizing a Bi-directional Long Short-Term Memory (Bi-LSTM) Recurrent Neural Network (RNN). The proposed model architecture includes a word-encoding layer to transform input text into word embeddings, a Bi-LSTM layer to capture context from both past and future inputs, and a softmax layer for classifying the dialogue acts. The use of word2vec for language modeling in MDAR enhances the system's ability to understand and process Myanmar dialogues more effectively.

A significant contribution of this work is the creation and annotation of the MmTravel corpus, which consists of 80,000 utterances from human-human travel domain conversations. The construction of the MmTravel corpus is especially crucial for low-resource languages like Myanmar, providing a robust data foundation necessary for training effective machine learning models. This corpus not only facilitates the development of the MDAR system but also contributes valuable resources to the broader NLP community, promoting further research and development in underrepresented languages.

The research reports a detailed analysis and comparison of the proposed Bi-LSTM model with traditional RNN, LSTM, and baseline SVM models. Experimental

results demonstrate that the Bi-LSTM model outperforms previous approaches, achieving an accuracy improvement of over 2% compared to the SVM model on the MmTravel corpus. This research not only advances in Myanmar dialogue act recognition but also contributes to the broader field of multilingual NLP by providing robust methodologies and resources for underrepresented languages. The insights gained from this research can be applied to other low-resource languages, paving the way for more inclusive and diverse NLP technologies.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1

INTRODUCTION

Dialogue Act Recognition (DAR) is a critical component in the field of natural language processing (NLP), playing a pivotal role in the interpretation and generation of human language. DAR involves identifying the function of a segment of dialogue, which is essential for various applications such as chatbots, virtual assistants, and automated customer service systems. Popular languages like English, Chinese, and Japanese have extensive resources and research dedicated to them, with large annotated corpora such as the Switchboard Dialog Act Markup in Several Layers (SWBD-DAMSL) corpus, which includes over 1,155 one-on-one five-minute telephonic conversations annotated into 42 different dialogue acts, and the ICSI Meeting Recorder Dialogue Act (MRDA) corpus, which contains 75 multi-party meetings labeled into more than 50 different dialogue acts. These resources have significantly advanced the development of DAR for these languages.

However, approximately seven thousand languages around the world, including Myanmar, suffer from a scarcity of data and resources, making DAR particularly challenging for these low-resource languages. In the context of Myanmar, a country with a rich linguistic heritage and a unique script, the development of Myanmar Dialogue Act Recognition (MDAR) is both a challenging and essential endeavor. Myanmar's linguistic landscape is characterized by its own set of syntactic rules, semantic structures, and tonal variations, which add layers of complexity to the task of dialogue act recognition. Myanmar's script, derived from the Brahmi script, includes characters and symbols that represent unique phonetic and grammatical elements, posing significant challenges for standard NLP tools designed for more commonly used languages.

MDAR aims to classify dialogue utterances into predefined categories, such as questions, statements, requests, and commands, to facilitate more effective human-computer interactions in the Myanmar language. This classification is crucial for understanding user intentions and generating appropriate responses, which is particularly important in multilingual and culturally diverse settings. To address these challenges, MDAR necessitates the development of sophisticated models capable of

capturing the nuanced linguistic features of Myanmar. Advanced machine learning techniques, particularly Bi-directional Long Short-Term Memory Recurrent Neural Networks (Bi-LSTM RNNs), are leveraged to enhance the accuracy and reliability of dialogue act classification. These models are well-suited for processing sequential data and can effectively handle the context-dependent nature of dialogue. Bi-LSTM RNNs can analyze the context before and after a given word or phrase, making them particularly effective for understanding the flow and structure of conversations in Myanmar.

By leveraging these advanced techniques, MDAR seeks to improve the performance of dialogue act recognition systems, thereby contributing to the broader goal of enhancing communication technologies in Myanmar. The development of reliable MDAR systems holds significant potential for various applications, including language translation services, educational tools, and accessibility technologies for individuals with disabilities. Moreover, as Myanmar continues to integrate more digital and automated services, the role of MDAR in facilitating smooth and intuitive interactions between humans and machines becomes increasingly vital. Overall, the progress in MDAR not only represents a technical achievement but also a step towards preserving and promoting the linguistic diversity of Myanmar in the digital age.

1.1 Motivation of the Thesis

The motivation behind this thesis stems from the growing importance of NLP applications in everyday life and the relative paucity of resources and research focused on the Myanmar language. Natural Language Processing (NLP) has become integral to numerous facets of modern life, powering technologies that facilitate seamless human-computer interactions. As digital platforms proliferate and automated systems are increasingly relied upon for customer service, education, entertainment, and more, the demand for robust NLP systems capable of understanding and responding to users in their native languages has never been greater.

Myanmar, despite being spoken by millions, lacks the extensive computational resources and research attention afforded to more widely spoken languages. This disparity is evident in the limited availability of linguistic data, computational tools, and language-specific models tailored to the unique characteristics of the Myanmar

language. Consequently, many NLP applications in Myanmar are either non-existent or fail to meet the same standards of accuracy and efficiency seen in languages with more research investment. This gap in resources and research presents a significant opportunity to develop innovative solutions specifically tailored to the linguistic characteristics of Myanmar.

By focusing on the development of Myanmar Dialogue Act Recognition (MDAR), this research aims to address the challenges inherent in understanding and processing Myanmar language dialogues. The goal is to create systems that can accurately classify dialogue utterances into categories such as questions, statements, requests, and commands, thereby enabling more effective human-computer interactions. Achieving this requires tackling the unique syntactic, semantic, and phonetic features of the Myanmar language, which differ significantly from those of more widely studied languages.

This research aims to contribute to the equitable advancement of NLP technologies, ensuring that speakers of all languages can benefit from the digital revolution. By developing sophisticated models and leveraging advanced machine learning techniques, such as Bi-directional Long Short-Term Memory Recurrent Neural Networks (Bi-LSTM RNNs), this thesis seeks to enhance the accuracy and reliability of NLP applications in Myanmar. These efforts will not only improve the user experience for Myanmar speakers but also serve as a blueprint for similar initiatives in other underrepresented languages.

Ultimately, the motivation driving this paper is to bridge the digital divide and promote linguistic diversity in the realm of NLP. By investing in the development of language technologies for Myanmar, this research hopes to pave the way for more inclusive and accessible digital solutions, ensuring that the benefits of technological advancements are shared equitably across different linguistic communities.

1.2 Objectives of the Thesis

The primary objective of this research is to develop a Bi-LSTM RNN-based Myanmar Dialogue Act Recognition (MDAR) system, with a focus on creating a robust framework for understanding and categorizing dialogue acts in the Myanmar language. To achieve this, the research begins by observing and analyzing Myanmar

Dialogue Acts to gain a comprehensive understanding of their unique characteristics. This foundational step involves examining the various types of dialogue utterances used in natural conversations in Myanmar. By systematically categorizing and documenting these dialogue acts, the research aims to establish a detailed overview of the interaction patterns and linguistic features specific to Myanmar, which is essential for the accurate modeling of dialogue acts.

A significant milestone in this research is the introduction of a Myanmar Dialogue Act Tagset for Dialogue Act Modeling. This tagset is meticulously designed to capture the diverse range of dialogue acts encountered in Myanmar language interactions. It provides a structured and standardized framework for tagging and annotating dialogue data, ensuring consistency and clarity in the subsequent stages of model training and evaluation. The creation of this tagset is crucial as it lays the groundwork for effective dialogue act recognition, enabling the Bi-LSTM RNN model to accurately classify and interpret various dialogue utterances based on their functional roles in communication.

Exploring different methods and possibilities for evaluating Myanmar Dialogue Act Recognition is another core objective of this dissertation. The research employs a comprehensive evaluation strategy, utilizing a variety of metrics and benchmarks to assess the performance of the Bi-LSTM RNN model. This includes traditional metrics such as precision, recall, and F1-score, as well as more nuanced evaluations that consider the specific challenges posed by the Myanmar language's syntactic and semantic intricacies. By comparing the proposed model's performance against existing methods, the research aims to highlight the advantages and limitations of the Bi-LSTM RNN approach, providing valuable insights into its effectiveness and potential areas for improvement.

The practical applications of the Bi-LSTM RNN-based MDAR system are vast and varied, extending the relevance and impact of this research beyond academic circles. One key application is in dialog systems, where accurate dialogue act recognition can enhance the system's ability to understand and respond to user inputs in a natural and contextually appropriate manner. This has significant implications for developing more intuitive and effective conversational agents in the Myanmar language. Additionally, the MDAR system can be applied to automatic question and

answering systems, where it can improve the capacity of the system to correctly interpret user queries and generate precise answers. In customer service, the implementation of MDAR can lead to more efficient and responsive automated support systems, capable of handling a wide range of customer inquiries in the Myanmar language with greater accuracy and efficiency.

In summary, the objectives of this research are multifaceted and aim to advance the field of Myanmar Dialogue Act Recognition through the development of a sophisticated Bi-LSTM RNN model. By observing Myanmar Dialogue Acts, introducing a comprehensive tagset, exploring various evaluation methods, and applying the MDAR system to practical applications, the research seeks to contribute to the equitable advancement of NLP technologies, ensuring that speakers of the Myanmar language can fully benefit from the digital revolution.

1.3 Contributions of the Thesis

This research makes several significant contributions to the field of natural language processing (NLP), particularly in the context of Myanmar Dialogue Act Recognition (MDAR) using Bi-directional Long Short-Term Memory (Bi-LSTM) Recurrent Neural Networks. One of the foremost contributions is the building of a large Myanmar Dialogue Corpus. This corpus, meticulously compiled and annotated, serves as a foundational resource for the research and development of MDAR systems. The Myanmar Dialogue Corpus is designed to capture the richness and diversity of spoken interactions in Myanmar, encompassing various dialogue types, contexts, and linguistic nuances. This extensive dataset not only facilitates the training and evaluation of MDAR models but also stands as a valuable resource for future research in Myanmar language processing, addressing the critical shortage of language-specific datasets in this underrepresented area.

In conjunction with the corpus development, the research proposes a comprehensive set of Myanmar Dialogue Act Tagsets. These tagsets are specifically tailored to the syntactic and semantic characteristics of the Myanmar language, providing a standardized framework for categorizing dialogue acts. By defining a clear and detailed taxonomy of dialogue acts, the research ensures consistent annotation and classification of dialogue data, which is essential for accurate model training and performance assessment. The introduction of these tagsets marks a

significant step forward in the formalization of MDAR, offering a robust tool for researchers and practitioners to systematically study and model dialogue interactions in Myanmar.

To establish a solid baseline for MDAR, the thesis explores various machine learning approaches, including the Naïve Bayes classifier and Support Vector Machine (SVM). These traditional classifiers are evaluated on the Myanmar Dialogue Corpus to provide a benchmark against which the performance of more advanced models can be measured. By investigating these baseline methods, the research offers valuable insights into their strengths and limitations in the context of Myanmar language processing. This comparative analysis not only highlights the challenges posed by the unique linguistic features of Myanmar but also underscores the necessity of employing more sophisticated techniques to achieve higher accuracy in dialogue act recognition.

Another pivotal contribution of this research is the application of word vector features on the proposed models. By leveraging word embeddings, which capture semantic relationships between words in a continuous vector space, the research enhances the ability of MDAR models to understand and interpret the context of dialogue utterances. Word vectors provide rich contextual information that traditional feature extraction methods might miss, thereby improving the overall performance of the dialogue act classification. The incorporation of word vector features represents a significant advancement in the modeling of Myanmar dialogues, aligning with state-of-the-art practices in NLP.

Central to the thesis is the development of a Recurrent Neural Network (RNN) model using Bi-directional Long Short-Term Memory (Bi-LSTM) for Myanmar Dialogue Act Recognition. The Bi-LSTM RNN architecture is particularly well-suited for sequence-to-sequence tasks like dialogue act recognition, as it can effectively capture long-range dependencies and contextual information from both past and future states of the dialogue sequence. This bidirectional approach allows the model to consider the full context of an utterance, leading to more accurate and nuanced classification of dialogue acts. The implementation and optimization of the Bi-LSTM model for Myanmar dialogues constitute a major technological advancement,

demonstrating its effectiveness in handling the complex syntactic and semantic structures of the Myanmar language.

In summary, the contributions of this research are multifaceted, encompassing the creation of a large and comprehensive Myanmar Dialogue Corpus, the proposal of detailed dialogue act tagsets, the exploration of baseline machine learning approaches, the application of word vector features, and the development of an advanced Bi-LSTM RNN model for MDAR. These efforts collectively advance the state of NLP for the Myanmar language, providing essential tools and methodologies that pave the way for more sophisticated and inclusive language technologies.

1.4 Organization of the Thesis

This research is meticulously organized to comprehensively explore dialogue act recognition in the Myanmar language. Chapter 2 reviews the Literature and Related Work, conducting a thorough examination of existing research, methodologies, and findings in dialogue act recognition. This critical analysis of prior studies identifies gaps and justifies further research. Chapter 3 focuses on the Construction of the MmTravel Corpus, a significant contribution of this dissertation. This corpus is meticulously designed to support the development of dialogue act recognition models tailored for Myanmar, detailing the processes of data collection, annotation, and validation to enhance linguistic resources for underrepresented languages.

Chapter 4 describes the machine learning-based Myanmar Dialogue Act Recognition, comparing algorithms like Support Vector Machines and Naive Bayes to determine their effectiveness. Chapter 5 introduces a Bi-LSTM RNN-based approach for Myanmar Dialogue Act Recognition, and exploring its architecture, training, and evaluation. This method highlights the improved accuracy in capturing contextual dependencies over traditional machine learning methods. Chapter 6 concludes with a summary of findings, contributions, and future research directions, underscoring the research's systematic progression and its implications for advancing dialogue act recognition, especially in lesser-represented languages like Myanmar.

CHAPTER 2

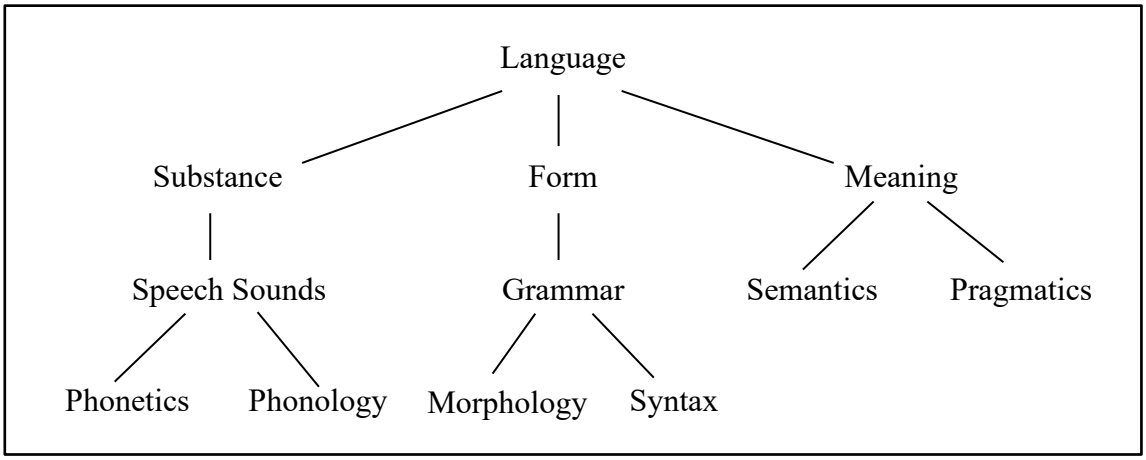
LITERATURE REVIEW AND RELATED WORKS

This chapter offers a comprehensive review of Dialogue Act Recognition techniques and related research, alongside essential background information on natural language processing (NLP) and dialogue systems necessary for understanding the rest of the study. Additionally, it briefly introduces the evaluation metrics used.

2.1 Linguistic Background

Myanmar (Burmese) language is a Tibeto-Burman subfamily of Sino-Tibetan languages which is spoken primarily in Myanmar. There are two complementary registers for Myanmar: colloquial style which is used in spoken conversations or informal writing and literary style which is used orally in formal setting. Speeches or news broadcasting are literary style [7]. This research concentrates on the facts for colloquial style.

Figure 2.1 Structure of the Linguistic



Every language, whether natural or artificial, follows a set of rules known as grammar that determines its proper usage. Only adequate knowledge of these rules allows one to properly understand and produce utterances in each language [20]. The rules that govern morphology and syntax are the most common application of the term grammar [83]. However, it can also encompass phonology, semantics, and pragmatics in a broader sense. Figure 2.1 shows the structural linguistic.

2.1.1 Morphology and Syntax

Morphology examines how words are constructed from smaller linguistic units, such as morphemes, including roots, stems, prefixes, and suffixes. Nouns like "ခရီးသည် (kʰəyí ðè, passenger)" get the suffix "-များ (myá)" in plural form because there are three general plural morphemes: "-များ (myà)", "-တို့ (tó/dó)" and "-တွေ (twe/dwe)" which are like the English plural "-s" in terms of its meaning. Adjectives such as "လှပ (l̰a̰ pa̰)" can be converted into adverbs by appending "-စွာ (swà)." Additionally, verbs can be modified in various ways based on tense, person, and number.

Syntax concerns the arrangement of words to create phrases and sentences. In the Myanmar (Burmese) language, which follows a head-final syntactic structure with a fundamental SOV (subject-object-verb) word order similar to Japanese, Korean, Turkish, and Hindi, a di-transitive clause can display various permutations of word order. While the verb-final property of such languages may be less strictly enforced, it is also possible for other combinations to occur, allowing arguments of the verb to be optionally placed after the verb.

The English language has extremely strict word order rules, while other languages are much more lenient. In English sentences, determiners (such as "the", "my", "this") must be placed before the noun they modify (e.g., "the cat ate the food"). Negation is accomplished by placing the term "not" immediately following specific verbs like "do", "can", and "be." In Myanmar language, it does not have the determiners and the negative morpheme "မ (ma̰)" is a particle prefixed to a verb to convey a negative sense "မ ... ဘူး (ma̰ ... bú)" used in declarative sentence may not be considered as part of modality. However, the morpheme used for negative imperative "မ ... နဲ့ (ma̰ ... bú)" conveys speech modality. Differentiating between morphology and syntax can sometimes be challenging. For example, the morphological features of a noun often rely on its syntactic connections with other words, as shown in the following sentence:

ခရီးသည်များ ကား ပေါ် တက် ပါ
(kʰəyíðè myà ká pɔ̀ teʔ pà, Passengers get in the car)

Here, the subject of the sentence "ခရီးသည် (kʰəyí ðè, passenger)" gets the suffix "-များ (myà)" because of the plural noun, which is relation to the verb "တက် (teʔ, get into)" and the object of the sentence "ကား (ká, car)". The term "morphosyntax" is used when both morphology and syntax are developed together.

2.1.2 Semantic

Semantics examines meaning across different levels, ranging from morphemes to complete sentences. It encompasses two primary subfields: lexical semantics, which explores the meaning of individual words and their interconnections. Lexical relationships include synonyms (words share the same meaning, such as "အမေ" (əmèi, mother) and "မိခင်" (mi̱ gĩ, mother)), antonymy (words with opposite meanings, such as "နီး" (ní, near) and "ဝေး" (wéi, far)), and hyponymy/hypernymy (a hierarchical relationship also known as IsA). In this hierarchy, a hyponym is a term that has a more specific meaning encompassed by its hypernym; for example, "အဖြူ" (əpʰyù, white) is a hyponym of "အရောင်" (əyaŏ̃, color).

Word meanings are accompanied by basic properties known as *semantic features*, which provide more information about how words relate to one another. "အဖြူ (əpʰyù, white)" and "အပြာ (əpyà, blue)" are two distinct colors, "ကြောင် (ʧaŏ̃, cat)" and "ခွေး (kwé̱i, dog)" are two different animals, and "ဒီနေ့ (dì ne̱i, today)" and "မနက်ဖြန် (maneʔ pʰyã̀, tomorrow)" are two separate times. These characteristics are significant as they determine which words can be semantically combined. Semantic features are particularly important for verbs because they limit the kinds of noun phrases that can serve as their arguments. For instance, the verb "ပြေး (pyéi, run)" necessitates one argument (a subject), which must be animate, as inanimate objects cannot run. To clarify the relationship between the verb and its arguments, these arguments can be assigned specific names, known as thematic roles. Examples include the agent, where the typically conscious participant performs the action; the theme, where the participant is affected by the action; and the recipient, where the participant receives something as a consequence of the action.

Compositional semantics explores how the meaning of a sentence arises from the meanings of its individual words. It uses set theory to assess the truth value of a sentence: each word or phrase is associated with a set, and the sentence is considered true if these sets intersect. For instance, take the following sentence:

ငှက် အဖြူရောင် တွန် နေ တယ်
(ŋ̊eʔ əpʰyùyaʊ̃ tṹ nèi tè, A white bird chirps.)

"အဖြူရောင် (əpʰyùyaʊ̃, white)" refers to the collection of all white things, "ငှက် (ŋ̊eʔ, bird)" to the collection of all birds, and "တွန် (tṹ, chirp)" to the collection of all chirping things. The intersection results in a group consisting solely of birds, each of which is white and chirping.

Compositional semantics struggles with certain situations, such as paradoxes, idioms, and anomalies. Paradoxes are self-contradictory and lack a clear truth value. Idioms have fixed meanings that can't be broken down. Anomalies are grammatically correct but lack meaningful interpretation, such as "အရောင်မဲ့ စိမ်းလန်းသော စိတ်ကူးများ ဒေါသတကြီး အိပ်ပျော်နေကြသည် (Colorless green ideas sleep furiously)." They are closely related to metaphors, which initially seem out of place but still convey meaning, for example, "အချိန် သည် ငွေ (əʤʰèĩ θè ŋwèi, time is money)."

2.1.3 Pragmatic

Semantics focuses on the literal interpretation of statements. However, the meaning of a sentence can vary based on the speaker, the listener, the timing within the conversation, and the intended purpose. In essence, the context plays a crucial role in shaping the meaning of an utterance. For example, "ဒီ မှာ အေး တယ် (dì m̥à éi tè, It's cold in here)" could be interpreted straightforwardly as an observation, a suggestion to increase the temperature, or even sarcastically implying that it's not cold. Pragmatics involves understanding what is meant when context is considered. Figure 2.2 illustrates the difference between the speaker's meaning and semantic meaning. Semantics focuses on the relationship between the lexicon, grammar, and semantic meaning. Pragmatics studies the relationship between the context of use and both semantic and speaker meaning [68].

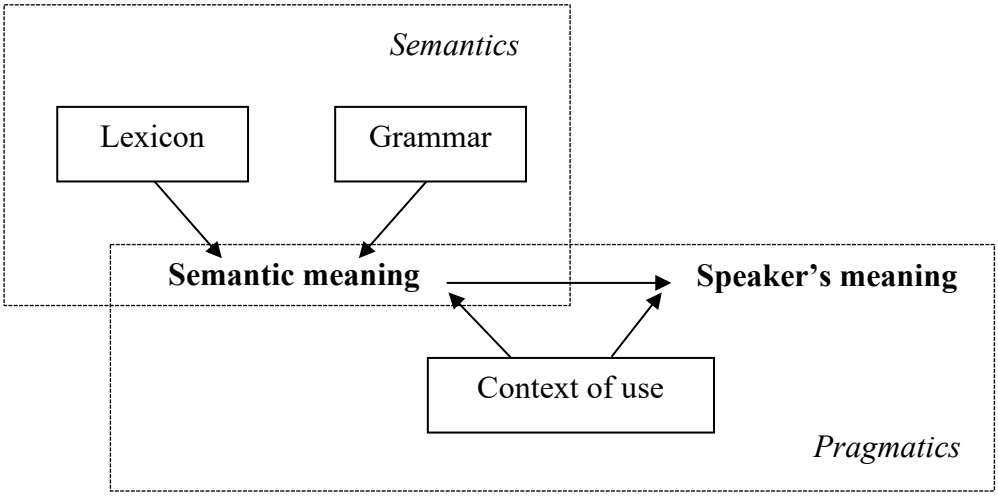


Figure 2.2 Speaker's meaning and Semantic meaning

Context can have a variety of effects on meaning. One of the most direct methods is deixis, which involves using words whose meanings change depending on the context. Deictic words include pronouns like "သူ (θù, he/she)" and "သူတို့ (θù to̰, they)"; demonstratives like "ဤ (ì, these)" and "ထို (tʰò, that)"; prepositions like "နောက်တွင် (naʊʔ twɪ̀, behind)" and "မတိုင်မီ (mə̰ tàĩ mì, before)"; and certain adverbs like "ဒီမှာ (dì m̥à, here)" and "မနက်ဖြန် (maneʔ pʰyɛ̃́, tomorrow)." While these words have some meaning on their own, context is needed to fully understand them. For instance, the word "သူ (θù, he/she)" refers to a person, but context is required to identify which person. Deixis exemplifies a broader approach to communication aimed at conveying messages quickly, easily, and socially safely. Consequently, speakers often leave much information unsaid, expecting the addressee to correctly fill in the gaps based on context. The act of suggesting something indirectly instead of stating it explicitly is referred to as implicature.

An important understanding in conversations is that speakers can accomplish actions through speech. These actions serve their communicative goals, often categorized as speech acts. In the field of computer science, these are more commonly known as dialogue acts, highlighting their role in computational linguistics and NLP research. This recognition underscores the pivotal role of language in achieving specific intentions and objectives within dialogues and communication contexts.

2.1.4 Speech Act Theory

When it comes to the question of how meaning and language are related, there are two major lines of thought in contemporary language philosophy. The logical positivist school of thought holds that the meaning of an utterance can be described by logical formulae that describe the facts to which the utterance refers. It focuses on the relationship between language and the physical world. It was founded by Frege and Russell, and Montague, Carnap, Tarski, Kripke, and others have further explored it. The other school of thought is known as speech act theory, based on language. Wittgenstein and Austin set it up, and Searle and Grice investigated it further. Whereas the first line assumes isolated propositions as the unit of meaning, the second assumes a combination of a proposition and how it is contextualized in language.

A sentence is meaningful only if it can be tested for truth or falsity, either in a purely formal logical framework (logical verification) or by observation, according to logical positivism (empirical verification). Unverifiable statements are regarded as nonsense. Logical positivism has made a significant contribution to the theory of sentence meaning by using logical formalisms as a foundation. The development of logics of sense — including temporal connectives and quantification — has made it possible to express aspects of propositional content. However, because a proposition-based account of natural language meaning is insufficient to account for certain contextual and performative uses of language, this section focuses on speech act theory.

Austin's ideas, as presented in [42], contrast with the assumptions advanced in a logical positivistic approach to language description. The basic type of language use is declarative, and the meaning of utterances can be described by propositional truth or falsity, according to these assumptions. Austin challenges these assumptions and makes two key points. To begin with, not all utterances are statements, and much of conversation consists of non-statements such as exclamations, questions, and expressions of desire, among other things. Second, not all utterances with grammatical forms that correspond to declaratives are used to make statements. According to Austin, these types of utterances have no truth-conditional statements and are actions in and of themselves. For example, when the speaker says, "မနက်ဖြန် မင်းကို ကူညီမယ်လို့ ကတိပေးတယ် (maneʔ pʰyɛ̃́ mḭ́kɔ̀ kù ɲì mè lo̰ gədḭ péi tè, I promise to help you tomorrow)", he or she is not just describing but also making a promise. This type of utterance was given the

name performative utterances or performatives, while other utterances were given the name constative utterances or constatives. Performative utterances have a distinct linguistic structure: they ‘perform’ the action specified by the main verb and are usually realized in the first-person present tense. This main verb belongs to a group of verbs that describe verbal activities such as promising, warning, and so on. Austin concludes later in his analysis that there is no theoretically sound way to distinguish between performatives and constatives, so he does away with the distinction by considering statements to be just another type of speech act, which he refers to as ‘stating’. As a result, Austin concluded that utterances are speech acts. The type of act is shown in some utterances by a main verb writing down the action performed; in others, the action is implied.

Austin distinguished between the speaker's intention and the effect on the listener when he describes an utterance as a type of action performed by the speaker. A speech act entails the expression of a meaningful message in a specific language (the locutionary act), the performance of an action (the illocutionary act), and the effects on a listener (the perlocutionary act). Asking, asserting, and promising are examples of illocutionary acts. Perlocutionary acts include things like persuasion and reassurance. The illocutionary act is the focus of Austin's research, as well as pragmatic studies on language use in general, and the terms 'speech act' and 'illocutionary act' are frequently used interchangeably. The view that a speech act can be decomposed into two parts: an illocutionary force and a propositional content is facilitated by the assumption that all utterances are speech acts that encode the meaning of an utterance (which need not to be realized). Consider the following three similar utterances and their accompanying speech acts:

ခရီးသည် တွေ ကား ပေါ် တက် သွား ပြီ (Inform)
(kʰəyíðè twèi ká pɔ̀ teʔ θwá pyì, Passengers got on the car.)

ခရီးသည် တွေ ကား ပေါ် တက် ပြီ လား (Question)
(kʰəyíðè twèi ká pɔ̀ teʔ pyì lá, Are the passengers get in the car?)

ခရီးသည် တွေ ကား ပေါ် တက် ပါ (Instruction)
(kʰəyíðè twèi ká pɔ̀ teʔ pà, Passengers, get in the car!)

The proposition “get (passengers, car)” can be used to describe the propositional content of each of the utterances listed above. However, the type of illocutionary force

used in each utterance is different: the first utterance is declarative, the second utterance is interrogative, and the third utterance is imperative. Austin investigated the space of illocutionary forces based on performative verbs in a language's vocabulary and proposed a set of classes to categorize them into. But this classification was insufficient, which lead Searle to propose a new, more precise classification.

Searle made a key point when he says that illocutionary acts frequently conflate with the verbs that convey them. As a result, while a taxonomy of illocutionary acts like the one proposed by Austin in his early work is useful, it is limited in applicability for several reasons. One significant limitation of such a taxonomy is that it would be language dependent. Furthermore, a taxonomy based solely on verbs may be insufficient to address all utterances that are not explicitly performative. To overcome these constraints, Searle proposes focusing on illocutionary forces rather than verbs, and introduces several classes of illocutionary force to characterize illocutionary acts [44] [45] [46] [47]. By emphasizing illocutionary force, Searle argues, like Austin, that illocutionary force is an essential aspect of meaning that cannot be explained by truth and falsity alone. A theory of action, rather than a theory of truth-conditional meaning, should be used to explain meaning because it is necessary to know what the addressee is to do with the proposition (if any) conveyed in an utterance.

The goal of speech act theory is to describe the effect of an utterance as such, but it frequently fails to express the meaning of the utterance in the context of the dialogue in which it occurs. This context is dynamic in an ongoing dialogue because beliefs, desires, and intentions change over time. Given that dialogue involves multiple participants, issues such as communication coordination and mutual understanding (i.e., grounding) must be considered. Dialogue, as a collaborative social activity, also involves the communication of joint action plans and social attitudes. The Searlean notion of speech act is incorporated in more recent work on dialogue modeling and spoken dialogue systems in a more elaborate type of act that can also address other conversational functions an utterance can play. These acts are frequently referred to as dialogue acts, and many dialogue act taxonomies include domain-specific information.

2.2 Dialogue System Architecture

A dialogue system is made up of three primary elements: the Natural Language Understanding (NLU) module, the Dialogue Manager (DM), and the Natural Language

Generation (NLG) module. At the heart of the system, the DM interprets user input, updates the system's state and relevant components, and chooses a suitable response. The NLU and NLG modules serve as intermediaries between human and computer languages. The NLU processes the natural language input to extract pertinent information and relays it to the DM, while the NLG takes the DM's response and translates it back into natural language. This section focuses on processing natural language input rather than generating output, so NLG will not be discussed further. Figure 2.3 provides an overview of a dialogue system and its components.

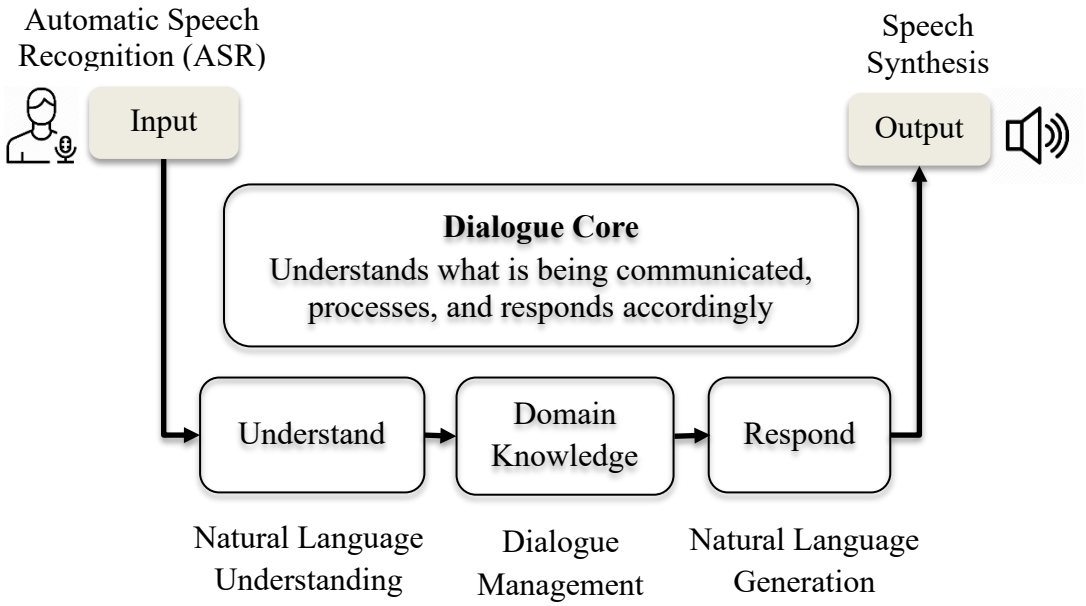


Figure 2.3 Overview of Dialogue System

2.2.1 Natural Language Understanding

In the case of text input, Natural Language Processing (NLP) aims to convert the unstructured output of Automatic Speech Recognition (ASR) into a structured text representation, encompassing Spoken Language Understanding (SLU) or Natural Language Understanding (NLU). At the time of writing, a common approach to NLU is to identify predefined keywords and phrases in the input. However, this method is not equivalent to true “understanding”; it is more like a cat recognizing a few commands. Despite this limitation, the method is effective for simple Dialogue Managers (DMs) such as finite-state or frame-based systems, which do not require much freedom of expression from the user. In contrast, information state and plan-based DMs require more sophisticated processing. These systems rely on dialogue acts as a core component and must at least identify the dialogue act of the input utterance.

Depending on the complexity of the conversations, they may also need to extract semantic information. Consequently, these systems require a more advanced NLU module capable of performing semantic and pragmatic analysis of the input.

The primary aim of semantic analysis is to determine the answers to questions like who performed an action, what was done, to whom it was done, when it happened, where it took place, why it occurred, and how it was carried out. These inquiries highlight semantic roles and their interconnections, usually signaled by verbs.

သူက သူမကို စာအုပ် တစ်အုပ် ရောင်းလိုက်တယ်
(θùkə̰ θùmə̰kò sàoʊʔ tiʔoʊʔ yáʊ̃laiʔtè, He sold her a book)

In the sentence, the verb "ရောင်း (yáʊ̃, sold)" serves as the relation, while "သူ (θù, he)", "သူမ (θùmə̰, her)", and "စာအုပ် (sàoʊʔ, book)" can be identified as the "seller", "buyer", and "goods", respectively. The method of assigning these roles to parts of a sentence is called semantic role labeling (SRL) or shallow semantic parsing. SRL typically involves four main steps. First, the target of the sentence, usually the main verb or predicate, is identified. Next, the meaning of the target within the context is determined, which defines the necessary semantic roles or arguments. Following this, all word sequences corresponding to each argument are identified. Finally, each argument is assigned an appropriate role label. [18] [54].

Pragmatic analysis aims to understand how the dialogue context affects the meaning of an utterance. It encompasses various subfields, such as deixis and conversational implicature, but a key area is dialogue act recognition. Dialogue acts are essential for both information state and plan-based dialogue managers. They enhance dialogue systems by simplifying the dialogue manager's tasks through abstraction and by restricting the system's response options. For example, despite varied user inputs requesting information, they can be managed similarly. Additionally, dialogue acts ensure appropriate responses, such as answering a question rather than replying with a greeting.

2.2.2 Dialogue Management

Dialogue management can be implemented in many ways, ranging from simple to complex. Each type has its own set of benefits and drawbacks, so the best option will

depend on the application. DMs can be based on four different types of models: finite-state, frame-based, information state, and plan-based [20].

2.2.2.1 Finite-state and Frame-based Models

Basic dialogue systems are typically categorized into finite-state and frame-based systems. Finite-state systems rely on predefined states and transitions, where the system produces responses based on its current state, and the user's input determines the transition to the next state. In contrast, frame-based systems operate by filling out forms, tracking parameters like travel dates and destinations. These systems ask the user questions to gather missing information, allowing questions to be asked in any order and accepting multiple pieces of information at once, thus offering more flexibility compared to finite-state systems.

In both types of systems, the dialogue is tightly controlled by the system, limiting the user's responses to a predetermined set of words and phrases. This results in straightforward, easily designed systems that do not require advanced technologies like speech recognition or natural language understanding. Dialogue act recognition is usually unnecessary since the system limits the user's actions to a specific set of options. However, these systems are quite restrictive and limited in their capabilities, and they can become cumbersome when handling more complex interactions, such as when users need to make revisions to their inputs. [56].

2.2.2.2 Plan-based Model

Plan-based methods to dialogue interpret conversations through the lens of planning, where utterances are seen as actions designed to fulfill particular communicative objectives. This perspective aligns with the concept of dialogue acts, where a series of actions can be chained together to form a plan to achieve a communicative objective. For instance, a travel planner might have a goal of providing users with the information they need. To achieve this, the planner could create a plan involving several “ask” actions to gather necessary details, followed by an “inform” action to present the final information. The ability to generate these plans dynamically sets plan-based approaches apart from other models. This allows for more complex conversations, such as those involving collaboration or negotiation, where the user can also take control. However, there are two main drawbacks. Firstly, allowing for more

complex language increases the difficulty of interpreting utterances. Secondly, systems based on plans often struggle with speed and complexity when addressing extensive problem areas. As a result, these systems are usually confined to small, specific domains that demand intricate dialogues. For applications requiring only straightforward interactions, simpler models are typically more suitable.

2.2.2.3 Information State Model

An information state model consists of an information state, dialogue moves, update rules, and an update strategy. The information state encompasses different types of data, including the dialogue history, the speakers' beliefs and intentions, and the appropriate type of response to provide next. Dialogue moves, akin to dialogue acts, are actions such as “ask” and “answer” that are carried out through natural language utterances and other forms of communication, like body language. When user input is received, the information state is modified by applying certain update rules. However, not every rule is applied in each instance; the dialogue moves identified from the user input determine which rules are relevant, and the update strategy decides which rules are activated and in what sequence.

The information state approach merges the benefits of finite-state and plan-based methods, offering greater flexibility by avoiding rigid, predetermined states and transitions [75]. It simplifies management and interpretation compared to plan-based systems, thanks to its constrained update rules and strategies. Moreover, this model adeptly handles complex information types, including beliefs, desires, and intentions, surpassing the capabilities of traditional plan-based approaches.

2.2.3 Natural Language Generation

The NLG task, in contrast to the NLU task, converts dialogue act concepts into sentences. It was first used in the 1990s for purposes like summarizing business and financial news. A few start-ups and large corporations also provide automated text generation solutions, which are used to quickly generate reports or official letters. The main challenge for such systems is to generate a wide and varied set of words, expressions, and sentences to avoid repetition and achieve the most realistic results possible. This is especially important when it comes to dialogue systems, which are meant to simulate real conversations with the user, which is a highly variable. Even

though data-driven solutions are becoming more mature, template-based generation methods are still the most popular in practice [28] [61] [27].

2.2.4 Dialogue System Modern Applications

Dialogue systems are used in a wide range of applications, from simple tasks such as controlling a home device or booking a flight to more complex tasks such as controlling a smart-room or managing traffic. Because of the complexity of managing language interfaces and their strong dependence on the interaction context, each application or set of applications necessitates the development of a unique model. As a result, most current prototyping techniques are limited to the creation of dialogue systems for a single or small group of applications. As a result, dialogue prototyping is crucial in the development of interactive systems, particularly those with a vocal interface: there is an urgent need for an efficient Rapid Dialogue Prototype Methodology [78]. The RDPM consists of five major steps:

- creating a task model for the application in question
- using the task model to generate an initial dialogue model
- experimenting with the Wizard-of-Oz to improve the initial dialogue model
- conducting an internal field test to improve the dialogue model (reformulation of system messages, improved feedback, etc.) and validate the evaluation procedure (coherence, understandability)
- using the evaluation procedure established during the internal field test, conduct an external field test to evaluate the final dialogue model.

Automatic dialogue act detection is used in a variety of DA applications. Dialogue systems, machine translation, Automatic Speech Recognition (ASR), topic identification, and talking head animation are the most important. DAs can be used to recognize the user's intent in dialogue systems, such as when the user requests information and waits for it, or when the system is attempting to interpret the user's feedback. ASR can use automatic detection of dialogue acts to improve word recognition accuracy. A talking head is a computer model of a human head that can accurately reproduce a speaker's speech in real time. It may also produce facial expressions that are relevant to the current state of the conversation. When a question

is asked, for example, using DA recognition in this context may improve the animation's naturalness by raising the brows. Another option is to use symbols and colors to display this additional information near the head.

Task-oriented conversational agents are designed to handle specific tasks efficiently through brief interactions with users, aimed at gathering necessary information to achieve task goals. These agents, such as digital assistants found on smartphones and home devices (like Siri, Cortana, Alexa, and Google Now/Home), can perform tasks such as providing directions, controlling home appliances, locating restaurants, and managing communications (e.g., calls and texts). Businesses employ goal-driven conversational agents on their websites to help customers find solutions to queries and resolve issues. Such agents serve as effective interfaces between humans and robots, with applications extending to beneficial social uses. For example, DoNotPay, a service dubbed a "robot lawyer," aids individuals in contesting parking tickets, applying for emergency housing, and seeking asylum, particularly beneficial for refugees.

Chatbots are computer programs designed to mimic the unstructured conversational or 'chats' characteristic of human-human interaction rather than focus on a specific task, such as booking a flight. Microsoft's 'XioaIce' system, which communicates with people via text messaging platforms, is an entertaining system. Attempts to pass various forms of the Turing test are frequently used by chatbots. Chatbots have been used for practical purposes such as testing psychological counseling theories since the first system, ELIZA [40]. It is worth noting that the term “chatbot” is frequently used as a synonym for “conversational agent” in the media and industry.

2.3 Dialogue Act Recognition (DAR)

The method of extracting meaning from natural language by determining the function of the text/sentence is known as *dialogue act recognition* (e.g., suggestion, question, command, or offer). In dialogue act recognition systems, a corpus in which sentences are labeled with the function, is used to train the model; and a statistical machine learning model that takes in a sentence and outputs its function is built. The model uses (i) words and phrases like “please” (function=request) and “are you” (function=yes/no question) as well as (ii) syntactic and semantic information to classify the sentences [56]. A labeled conversation is depicted in Table 2.1.

Table 2.1 Example of a Labeled Conversation (from the Switchboard corpus)

Speaker	Dialogue Act		Utterance
A	Conventional-opening	Fp	Hello?
B	Conventional-opening	Fp	Hi.
A	Repeat-phrase Statement-non-opinion	fp^m sd	Hi, My name is Leslie.
B	Repeat-phrase Statement-non-opinion	fp^m sd	Hi, I'm Jennifer.
A	Wh-Question	qw	Where are you from?
B	Statement-non-opinion	sd	Pennsylvania.
A	Appreciation Statement-non-opinion	ba sd	Nice, I'm from Dallas, Taxes.
B	Backchannel in question form	bh	Really?

Defining the relevant functions, referred to as the DA tag-set, is the initial step in creating a dialogue act recognition system. This process involves selecting labels that are broad enough to apply across various tasks, yet specific enough to remain pertinent to the specific task at hand, and distinct enough for humans to easily label the functions of sentences in the training dataset. Commonly used tag-sets in dialogue systems include Dialogue Act Markup in Several Layers (DAMSL), Switchboard SWBD-DAMSL, Meeting Recorder, VERBMOBIL, and Map-Task.

2.3.1 Dialogue Act Classes

A human conversation is a joint activity between two or more people. There always have a speaker and a listener in this activity. The conversation moves forward with the listener inferring about the speaker's speech and taking the opportunity to express his own ideas [20]. Austin [42] defines utterances as actions performed by the speaker. "Performative verbs" are verbs that specify actions, such as "I name this ship the Titanic." These are referred to as "speech acts." Speech acts, on the other hand, are not limited to these types of verbs. Searle [44] proposes five types of speech acts:

- **Representatives**: These are actions in which the speaker asserts something as true, including activities like suggesting, hypothesizing, insisting, swearing, concluding, and complaining.

- **Directives**: These actions involve the speaker trying to get the listener to do something, such as ordering, requesting, pleading, inviting, advising, defying, and challenging.
- **Commissive**: These are actions where the speaker commits to a future action, including promising, planning, threatening, swearing, vowing, and betting.
- **Expressive**: These actions convey the speaker's feelings or emotional state about a situation, such as thanking, congratulating, apologizing, condoling, and welcoming.
- **Declarations**: These are actions that bring about a change in the external world, such as resigning, nominating, firing, marrying, declaring, and convicting.

The intensity of representative, commissive, and expressive acts can differ significantly. For instance, insisting on something implies a stronger commitment than merely suggesting it. Likewise, issuing an order is more forceful than pleading, and making a vow is more substantial than making a plan. The verbs related to these categories can be versatile. For example, the verb "swear" can function in both a representative sense — "အမှန် ဆိုတာ ကျိန်ပြောရဲတယ် (əm̰ə̃ sʰò tà ʧə̀ĩ pjɔ́ yé tɛ̀, I swear it's true!)"—and a commissive sense — "အခဲ့ဒါ လုပ်မယ်လို့ ကျိန်ပြောရဲတယ် (ɛ̰ dà louʔ mɛ̀ lo̰ ʧə̀ĩ pjɔ́ yé tɛ̀, I swear I will do it!)"

The DAMSL tag-set offers a more thorough annotation system [57] compared to Searle's categorization [47]. Unlike Searle's approach, DAMSL permits an utterance to carry multiple labels, reflecting the reality that a single utterance can perform various actions. Additionally, DAMSL supports sub-classing, enabling systems to add custom classes as necessary while maintaining high comparability with other systems.

SWBD-DAMSL [8] is an adaptation of the original DAMSL framework, incorporating hundreds of potential dialogue act tags. For annotating the Switchboard corpus [37], 220 of these tags were utilized, which were then grouped to reduce the number of tiny classes, resulting in a more practical set of 42 dialogue act tags. These remaining classes are detailed in Appendix A. SWBD-DAMSL serves as a common base for various tag sets that are often modified or expanded as needed [71] [82]. The

classification of dialogue acts is a topic of debate, with some experts criticizing traditional taxonomies for relying more on intuition than empirical data [77]. This inconsistency can make it challenging for individuals to accurately tag utterances with the correct dialogue act. Consequently, some researchers have experimented with clustering algorithms to automatically identify categories, which were later given semantic labels by human evaluators [85].

2.3.2 Corpora

Dialogue act recognition can be thought of as a classification task in which each utterance is assigned a specific dialogue act label and its accuracy is measured. Multiple corpora with dialogue act annotations are available in terms of data.

Table 2.2 Annotated Corpora Characteristics

Corpus	Dialogue Act Tags	Utterances	Domain	Language
SWDA [37]	44	223,606	Open	English
ICSI-MRDA [[3]]	55	105,000	Meeting	English
HCRC Map Task [2]	12	26,621	Route Communication	English
NESPOLE [24]	1,168	12,565	Tourism	Multiple
VERMOBIL [59]	18 + 54	58,961	Appointment Scheduling	Multiple
SCHISMA [74]	64	440	Theatre Information	Dutch
AMI Meeting [34]	15	-	Meeting	English
LEGO [6]	22	5,188	Bus Information	English
Cambridge Restaurant [58]	13	932	Restaurant Information	English

The corpora and their characteristics are listed in Table 2.2. Multiple domains, languages, and types of interaction are covered, allowing portability experiments and domain and interaction independent conclusions. The used tag sets, on the other hand, are not standardized across corpora. The tag sets range in size from 12 (HCRC Map Task Corpus) to 1168 (NESPOLE) tags. Furthermore, while some tag sets are domain

independent (e.g., Switchboard Dialog Act Corpus [37]) and thus can be used to annotate any corpus, others (e.g., LEGO) are domain specific and thus limited to corpora in that domain. This means that the tag sets were created with different goals in mind and have different hierarchies and levels of abstraction, making inter-corpora experiments difficult to carry out. Furthermore, while some corpora (e.g. Switchboard Dialog Act Corpus, ICSI-MRDA [3] and HCRC Map Task Corpus [2]) have a large number of annotated utterances where most studies use one of these corpora to save time, but some choose to annotate their own corpus that better meets their needs with a small number of annotated utterances (e.g. SCHISMA and Cambridge Restaurant Corpus).

2.3.2.1 Switchboard Corpus

The DAMSL annotation scheme classifies sentences along four dimensions: communicative status, information level, forward-looking function, and backward-looking function. Communicative status identifies whether a sentence is unintelligible, abandoned, or self-directed speech. Information level categorizes a sentence as related to the task, task management, communication management, or another category. Forward-looking functions, divided into eight sub-dimensions, represent the sentence's impact on the future conversation: (i) assert, reassert, or another type of statement, (ii) influencing the addressee's future actions through options or directives, (iii) requesting information, (iv) committing the speaker to future actions via offers or commitments, (v) traditional openings or closings, (vi) explicit performatives, (vii) exclamations, or (viii) other. Backward-looking functions, such as agreement, understanding, responses, or informational relationships, show the connection between the current and preceding speech.

Switchboard [37] is a diverse dataset of informal phone conversations between two individuals, covering a wide array of topics. Participants were only required to have a casual chat rather than accomplish a specific task, resulting in largely unrestricted speech that encompasses various dialogue acts. The dataset is available in both original audio recordings and transcriptions. Over three months, eight Linguistics graduate students from the University of Colorado Boulder (CU-Boulder) annotated these dialogues with dialogue acts. The corpus includes 1155 dialogues (205K utterances, 1.4M words) annotated with 220 tags from the SWBD-DAMSL tag set [8].

Additionally, these tags have been clustered, reducing the number to forty-two by merging smaller classes.

The highest reported accuracy for the Switchboard corpus is 89.2% [4], though this study only used four types of dialogue acts (Incomplete, Statement, Question, and Backchannel), simplifying the problem. Another study achieved an accuracy of 80.7% after clustering the tag set [64]. The best result for the entire clustered tag set is 77.85% [9].

2.3.2.2 ICSI Meeting Corpus

The ICSI Meeting corpus [3] comprises 75 meetings with an approximate total of 795,000 words, involving ICSI working teams averaging six members each. The conversations are somewhat structured due to agenda items, yet they maintain a casual nature, similar to those in the Switchboard corpus. Both audio recordings and transcripts are provided. The corpus is annotated using the MRDA tag set [71], which is a modification of SWBD-DAMSL, featuring eleven general tags and thirty-nine specific tags. Each utterance is assigned one or more general tags along with a varying number of specific tags.

The highest accuracy reported for this corpus is 89.2%, but this was achieved using only five dialogue acts, not the full tag set [22]. Another study using just the eleven general tags achieved an 80.5% accuracy [63], while a different approach that clustered the complete set into sixty-two tags resulted in a 66% accuracy [36].

2.3.2.3 HCRC Map Task Corpus

The HCRC Map Task Corpus (Anderson et al., 1991) contains 128 audio recordings of pairs working together to achieve a specific objective. Consequently, the dialogues are task-focused, leading to more structured and formal language compared to casual conversations. Transcriptions with various annotations, including dialogue acts, are provided.

The corpus has been tagged using a set of twelve dialogue acts [34] that are mainly centered on the task, such as Instruct, Explain, Query-YN (yes-no question), and Reply-Y (yes answer). Therefore, it lacks categories for more social functions like

greetings and apologies. The highest recorded accuracy for this corpus is 73.91 percent [70]. In recent years, its use has declined significantly.

2.3.3 Rules vs. Statistical Methods

Natural Language Processing (NLP) can be broadly categorized into two main approaches: rule-based and statistical methods. Initially, NLP research predominantly employed rule-based techniques. However, due to the limitations of this approach and the rising prominence of machine learning, there was a significant shift towards statistical methods in the late 1980s, which continues today.

Rule-based systems [23] for language processing depend on manually crafted rules, which function similarly to if-then statements: a specific condition triggers a corresponding action. For example, if the system detects the word "hello," it might respond with "hi" or another greeting. These systems are straightforward to manage, debug, and comprehend, but as the number of rules increases, maintaining them becomes more challenging. Additionally, defining rules for complex scenarios can be problematic, as certain cases or exceptions might be overlooked, and rules could conflict with one another [38]. Crafting all necessary rules manually is often a labor-intensive task requiring specialized knowledge.

Machine learning techniques are employed in statistical methods to autonomously discern the general probabilistic patterns inherent in natural language. Given a sufficiently large and representative dataset for training, this approach is significantly quicker and less labor-intensive compared to manually creating rules. Additionally, statistical methods excel at handling grammatically incorrect or unusually phrased expressions, which are typical in human dialogue [65]. However, a major drawback is that achieving reliable results necessitates a vast amount of training data, which can be challenging to obtain. Moreover, the resulting models are primarily numerical and thus hard for humans to interpret. [52].

Various classification algorithms have been employed for dialogue act recognition, as shown in Table 2.3. Support vector machines (SVMs) are the most commonly used, often combined with hidden Markov models (HMMs) to create a hybrid system. Other frequently used methods include conditional random fields (CRFs) and Naive Bayes. Techniques such as Bayesian networks, maximum entropy,

logistic regression, and decision trees are less common. Interestingly, despite the recent popularity of neural networks in machine learning, they have not yet been applied to this task. This could be due to the relatively lower interest in dialogue act recognition compared to fields like computer vision.

The outcomes of classification algorithms cannot be accurately compared unless they are trained and tested in identical environments. This is a challenge because most studies only utilize a single classifier. However, some researchers have experimented with multiple classifiers. Two studies concluded that decision trees surpass Naive Bayes (Moldovan, 2011; Samei et al., 2014), while another study found decision trees to be superior to both SVM-HMMs and CRFs (Kim, 2010). In that same study, CRFs outperformed SVM-HMMs overall. Nonetheless, CRFs performed worse than SVM-HMMs in one scenario but better than standard SVMs in another (Tavafi, 2013). The conditions were not completely equivalent in this latter case, as the CRF considered the dialogue structure (such as the sequence of dialogue acts), unlike the SVM.

Gambäck et al. [9] used SVMs in conjunction with an active learning approach to perform dialogue act recognition on the Switchboard Dialog Act and Dihana corpora. Multiple n-grams, punctuation, and wh-words were among the features used. Using this method, the authors achieved 94.08% on the Dihana corpus [43] with seventy-two categories and 90.97% with 248 categories. On the Switchboard Dialog Act Corpus, the achieved accuracy results were 76.50%, 76.34%, and 77.85%, depending on whether the tag set had 42, 43, or 44 categories, respectively.

Grau et al. [72] combined the Naive Bayes classifier with the bag-of-words method: the feature set is made up of binary features, each of which indicates the presence or absence of a specific word. They also used a modified version of the classifier (uniform Naive Bayes), which ignores the DA probability. They have a DAMSL-switchboard corpus and a DA taxonomy score of 66%. The author of [Ivanovic, 2005] described applying the same modification to a subset of the DAMSL tag set (12 tags) and achieving 80% precision. Levin et al. [53] use the bag-of-words technique as well, but instead of word features, they used binary grammatical features. They achieved a precision of 51% using the NESPOLE corpus.

2.3.4 Deep Learning Approaches to DA Classification

A significant milestone in NLP research was paved by finding that the words can be efficiently represented as vectors. Words could be represented as vectors on a multidimensional space of word embeddings, according to Mikolov et al. [80]. Though Rumelhart et al. [19] studied the idea of having a distributed representation of words years ago, this research provided insights on how embeddings of words with similar meanings are clustered close to one another and how the distance vector between them indicates their contextual relation. Most researchers working on the DA classification problem had previously treated words as atomic units, but instead chose to represent them by enumerating the words and assigning an integer index to each one. Because of the contextual relationship between representations of each word, having a multidimensional representation for each atomic unit of utterances in a dialogue helped the methods that learned to extract implicit features in an automated way. The author [51] used the word vectors to train a convolutional neural network for sentence classification. The findings of that study show how important pre-training word vectors are in learning-based approaches to NLP problems.

Machine learning methods are used in most current DA classification research. Lee and Dernoncourt et al. [41] used recurrent neural networks (RNN) and convolutional neural networks to model short-text classification (CNN). This task is divided into two parts in their models. The first part, which uses either the RNN or CNN architecture, generates a vector representation of entire utterances based on the vector representation of words discussed above. The current utterance is then classified by considering its vector representation as well as a few previous utterances. With the widely used datasets SwDA and MRDA, which were also mentioned above, their method produced state-of-the-art results.

Kumar et al. [31] developed a model based on Graves' and Schmidhuber et al. [7] 's bidirectional Long Short-Term Memory (LSTM) units. Their model uses two layers of bidirectional LSTM units and has access to the entire conversation. The first layer creates an utterance representation, while the second layer considers all utterance representations in a conversation. The classification is then done by a CRF layer on top. With the SwDA dataset, this model achieved a near-human level of DA classification,

considering that its results were only 5% lower than the dataset's 84% inter-annotator agreement.

For the DA Classification problem, Bothe et al. [13] proposed a character-level RNN model. Krause et al. [10] investigated a multiplicative LSTM network as a model. Importantly, when the model considers an utterance in a dialogue, it only has access to previous and current utterances in that dialogue, but not future ones, to make the proposed model applicable to real-world human-computer interaction scenarios.

2.3.5 Evaluation for Dialogue Act Recognition

When evaluating a classifier's performance, precision and recall are key metrics calculated for each label. *Precision* measures the percentage of utterances correctly tagged as a specific dialogue act. For instance, a classifier always assigning the majority class will have 100% precision for that class but 0% for others, highlighting a lack of overall precision. Equation 2.1 gives the precision, where fp is the number of false positives and tp is the number of true positives. This indicates how frequently a positive guess for a specific label is correct.

The fraction of all utterances correctly classified as belonging to a specific dialogue act. It demonstrates how well a dialogue act is recognized. Equation 2.2 gives the recall, where fn is the false negative. Recall gives how many instances of the label it finds. The harmonic mean of precision and recall is the F-score. It is simply a method of combining precision and recall into a single value, and it can be thought of as a kind of classifier accuracy metric. Equation 2.3 can then be used to calculate the F1-score using recall and precision.

$$P = \frac{tp}{tp + fp} \tag{2.1}$$

$$R = \frac{tp}{tp + fn} \tag{2.2}$$

$$F = 2 . \frac{P . R}{P + R} \tag{2.3}$$

$$A = \frac{c}{G} \tag{2.4}$$

It is also preferable to assess performance for an entire guess rather than per label. The metric accuracy will be used for this purpose. *Accuracy* is the ratio of the

number of correct predictions true positives (tp) and true negatives (tn) to the total number of predictions correct predictions and false positives (fp) and negatives (fn), because it is consistently chosen as the measure to evaluate performance in dialogue act recognition. Equation 2.4 gives the accuracy, where G is the total number of guesses and c is the number of guesses with the correct value for each label. If a single label in the guess has the incorrect value, the guess is considered incorrect.

Precision, recall, and F-score for each dialogue act are calculated separately and then averaged using both micro and macro averages. The macro average gives equal weight to each class, while the micro average is weighted by class size. In multi-class scenarios, the average recall equals the average accuracy. The micro average can be skewed by larger classes, whereas the macro average provides a balanced performance view across all dialogue acts. Due to the lack of fully disclosed training and testing partitions in the corpora, a comparative study isn't possible. Instead, 10-fold cross-validation was used to reduce outlier influence and obtain a generalized view. The final precision, recall, F-score, and confusion matrix were calculated by averaging the results from the individual folds.

2.4 Summary

This chapter provides a brief introduction to background information on characterizing communication as action (speech act theory) because dialogue act tag annotation requires linguistic knowledge. It also expresses a variety of the most well-known annotated English corpora, and their dialogue act classes. The differences between rule-based and statistical approaches are also examined. The various DAR techniques are reviewed, and the experimental results of each approach are also presented. Deep neural networks have received attention in recent years due to their ability to reduce feature engineering, and the deep learning approach has become the preferred method of many researchers. Furthermore, model evaluation is covered in this chapter.

According to a review of previous research on Myanmar language, there is no research on DAR for Myanmar language that uses not only statistical methods but also neural network-based architectures.

CHAPTER 3

CONSTRUCTION OF AN MmTravel CORPUS

The construction of the Myanmar Travel (MmTravel) corpus is a significant contribution of this dissertation, essential for dialogue act recognition in the Myanmar language, specifically for travel-related conversations. It provides a domain-specific dataset that allows the recognition system to accurately understand and classify intents and actions in travel dialogues, such as booking inquiries or itinerary changes. This targeted corpus addresses the unique linguistic and cultural aspects of Myanmar, often missed in generalized datasets. By offering a contextually rich dataset, the MmTravel corpus improves the accuracy and effectiveness of dialogue systems, enhancing the user experience in travel applications.

The MmTravel corpus is the foundational dataset for machine learning and deep learning-based dialogue act recognition in the Myanmar language. This corpus serves as the initial step in the process, providing the essential data needed to train and evaluate models. By analyzing and labeling the dialogues within this corpus, models can learn to identify and categorize various dialogue acts, facilitating the development of more sophisticated and accurate dialogue systems.

This chapter describes the process used to collect and prepare data for the MmTravel corpus used in this study. The development of a dialogue act tag set, which is used to categorize utterances, is described. The MmTravel corpus was created, and the annotation scheme was used to mark up the corpus with dialogue acts.

3.1 General Background of Myanmar Language

Myanmar, formerly known as Burma, has Myanmar (Burmese) as its official national language. Since confusion about the use of the two names, primarily among English speakers. However apart from political reasons, the name change reflects an important aspect of the country's language: Myanmar (Burmese) distinguishes between "colloquial" and "literary" language usage. The word "Burma" of "Burmese" is colloquial usage, and "Myanmar" used in literary usage. Both are typically used in conjunction with another word. The word "ဗမာ (ba-ma)" or "မြန်မာ (mya-ma)" is

followed by the lexical item “spoken language” or “written language” to refer to a language; for example, “မြန်မာ စကား (mya mà zə gá)” and “ဗမာ စကား (ba mar zə gá)” for spoken Myanmar, and ″ဗမာ စာ (ba mar sà)″ and ″မြန်မာ စာ (mya mà sà)″ for written Myanmar.

Other terms for "colloquial" and "literary" Burmese are often equated with "informal" and "formal" or "spoken" and "written" Burmese. While the distinction between spoken and written Burmese relates to language usage style, the difference between colloquial and literary Burmese pertains to the medium of text delivery. Linguists find it challenging to distinctly categorize spoken and written discourse, as these forms frequently overlap. Contrary to common assumptions, colloquial Burmese is not solely for speaking, nor is literary Burmese exclusively for writing. For instance, radio broadcasters may use the literary style for news announcements, while personal letters can be colloquial, and fiction often blends colloquial dialogue with literary narration. The main difference between the two styles is vocabulary: one form of a word is used in colloquial, while another is used in literary. For example, in colloquial, the word for “name” is “နာမည် (nə̃ me, name)” and, which is used “အမည် (əmyìn, name)” in literary. The example sentences of literary and colloquial language were expressed in below:

Colloquial Style:	ဆုဆု စစ်တွေ သွားတယ် (sʰṵsʰṵ siʔ twèi θwátè)
Literary Style:	ဆုဆု သည် စစ်တွေမြို့ ကို သွားသည် (sʰṵsʰṵ θè siʔ twèi myo̰ kò θwátè)
Sentence:	Su Su goes to Sittwe.
Colloquial Style:	မြန်မာနိုင်ငံ သွားပြီး ဗုဒ္ဓဘာသာ လေ့လာတယ် (myamà nàĩ ŋə̃ θwá pyí boʊʔ dạ̀ bà ðà lẹ̀i là tè)
Literary Style:	မြန်မာနိုင်ငံသို့ သွား၍ ဗုဒ္ဓဘာသာ လေ့လာ၏ (myamà nàĩ ŋə̃ θo̰ θwá ywẹ̀i boʊʔ dạ̀ bà ðà lẹ̀i là i.)
Sentence:	He traveled to Burma and studied Buddhism.

3.2 Myanmar Speech Functions

Sociology studies society, while linguistics focuses on language elements like phonemes, morphemes, and sentences. Language serves as a communication tool for people, who use various forms in different social contexts, shedding light on language function, community relationships, and social identity construction. According to Holmes (2013), a functional approach to language examines how language is used to serve different purposes through speaking, listening, reading, and writing [30]. The term "function" here refers to the roles different words or phrases play in communication, not their grammatical roles. Understanding these functions is crucial for effective communication, as they help listeners understand the speaker's intent.

The textual function of language connects language to itself and the context, making discourse possible by allowing speakers or writers to create recognizable texts. This function involves not only the relationships between sentences but also the internal organization of sentences and their meaning within and outside the context. Texts, whether spoken or written, long or short, operate as units of language. Sociolinguistics helps clarify these aspects by examining the relationship between language and society, enhancing our understanding of language structure and function in communication. Ultimately, language conveys information and expresses social relationships, establishing or confirming social identities and relationships while delivering the intended message.

Nowadays, conversation has become an important means of exchanging ideas in modern society, and it cannot be separated from human life. When attempting to express themselves, people not only produce grammatical structures and words, but they also perform actions through those utterances. For examples, "မင်း ကို အလုပ် ထုတ် လိုက် ပြီ, (mǐ kò əlouʔ htou' lai' pyì, You're fired.)", in this utterance, if you work for a powerful boss, the use of the expression means more than just a statement. The boss uses the phrase to perform the act of losing your job. Alternatively, the actions taken by utterances do not have to be as dramatic or unpleasant. Speech acts are actions performed through utterances and are labeled in Myanmar and English with more specific labels such as apology, complaint, compliment, invitation, promise, or request. These descriptive terms for various types of speech acts describe the speaker's

communicative intention in producing an utterance. The speaker usually assumes that the listener will understand what he or she is saying. The context of the utterance usually helps both the speaker and the listener in this process. The basics of speaker and listener communication can be found in the Figure 3.1. The addresser, also known as the speaker, writer, or sender, is the person who creates the subject matter. The person to whom the messages are addressed or sent is called the addressee, also known as the hearer and reader. A contribution in a dialogue has an addresser, who makes the contribution subject matter via communication channel, and an addressee, who is intended to receive and to process the contribution.

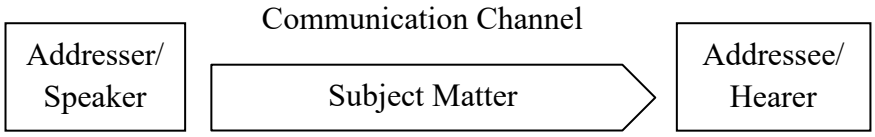


Figure 3.1 Addresser and Addressee Communication Channel

There is more to interpreting a speech act than what is contained in the utterance alone. On any given occasion, the act of producing an utterance is composed of three distinct acts. A *locutionary act* denotes the basic act of utterance or producing a meaningful linguistic expression. You may be unable to perform a locutionary act if you have difficulty forming the sounds and words required to make a meaningful utterance in a language.

Utterance_1: ကော်ဖီ နည်းနည်း ဖျော်ထားတယ်
(kɔ̀ pʰì né né pʰyɔ̀ tʰá tè, I've just made some coffee.)

Well-formed utterances are not usually made simply for the sake of making them. An utterance is created with a specific function in mind. This second dimension, known as the *illocutionary act*, involves using the communicative force of an utterance to perform tasks. Utterance_1 can be used to make a statement, extend an offer, provide an explanation, or communicate in other ways. This is also recognized as the illocutionary force of the utterance. Naturally, a functional utterance is not created unless there is an intention for it to have an effect. The third dimension is the perlocutionary act. Depending on the situation, Utterance_1 is uttered with the expectation that the listener will understand the desired effect (for instance, to express admiration for a pleasant aroma, or to prompt the listener to have some coffee). This is

referred to as the perlocutionary effect. Figure 3.2 illustrates the differentiation among the locutionary, illocutionary, and perlocutionary acts of a simple utterance.

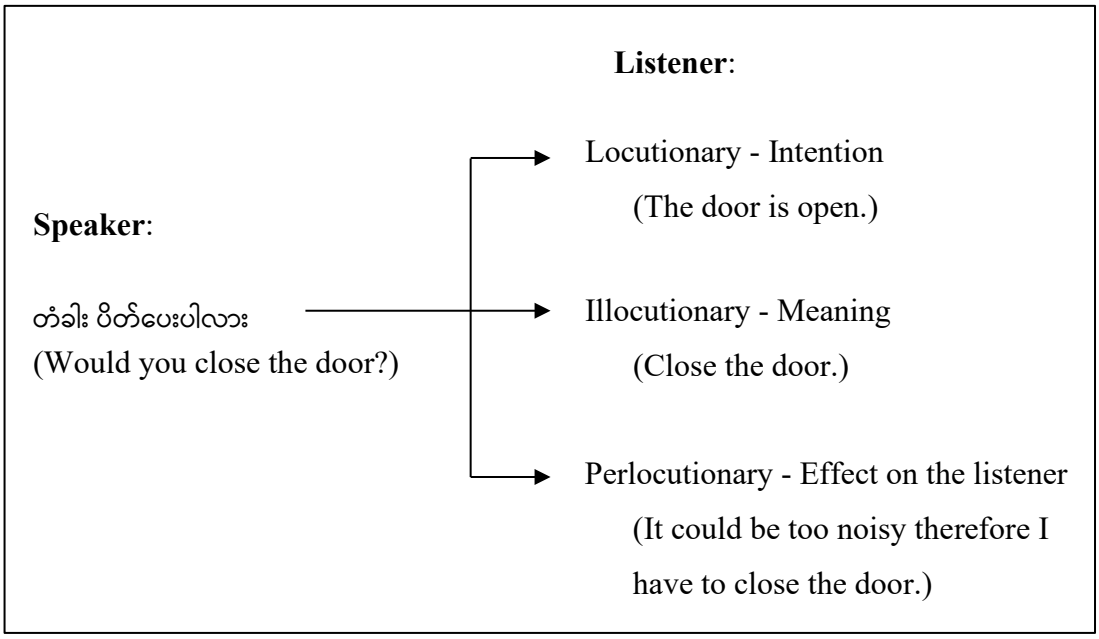


Figure 3.2 Locutionary, Illocutionary, and Perlocutionary Act

One general classification system classifies speech acts into five types of general functions, which are informational, expressive, directive, phatic, and aesthetic language function.

3.2.1 Informational Function

Everyone assumes that the most important function is the informational function. This function, in fact, focuses on the message. It is used to disseminate new information. It is determined by truth and value. The purpose of information is to persuade the speaker (to varying degrees) that something is true, that the stated proposition is true. An information, a description, a claim, a statement of fact, a report, and a conclusion are examples of worldly states or events. As a result, evaluating an informative is as simple as determining whether it is true or false. In the informative function, speakers and writers use language to express what they know or believe, which is concerned with facts. The purpose is to inform. By using informational language, the speaker makes words fit the world (of belief). For example,

Utterance_2: ကမ္ဘာကြီးက ပြားတယ်
(gəbà kji: kạ̰ pyá tè, The earth is flat.)

Utterance_3: မိုးရွာ နေ တယ်
(mó ywà nèi tè, It is raining.)

The speaker believes that the two utterances represent the world's events. The speaker states in Utterance_1 that he or she believes the earth is flat. The speaker's information about the current weather situation is implied by Utterance_2.

3.2.2 Expressive Function

The ability of a speaker to express his or her emotions is referred to as expressive function. A speaker addresses a message, and this focuses on the addresser. The goal of the expressive function is to convey the speaker's emotion, feelings or expression, and attitudes of the addresser. It is associated with the addresser and is also referred to as the emotive function. The speaker's intention is to give a direct expression of his or her feelings about the subject. It tends to give the impression of a specific emotion, whether truthful or not. It means that the attitude of the addressee toward the message's content is emphasized. For example,

Utterance_4: တောင်းပန် ပါ တယ်
(táõ bə̃ pà tè, I'm really sorry!)

Utterance_5: ဂုဏ်ယူ ပါ တယ်
(gòõ yù pà tè, Congratulations!)

This feature allows you to express personal feelings, thoughts, ideas, and opinions using different words, intonation, and other techniques. These expressions are influenced by social factors and the nature of the expression, which can be negative or positive.

Negative Utterance: အရမ်း စိတ်ဓာတ် ကျ နေ တယ်
(əyə̃ eiʔ daʔ ʧa̰ nèi tè, I'm very depressed.)

Positive Utterance: အရမ်း ပျော် တယ်
(əyə̃ pyɔ̀ tè, I'm very happy.)

3.2.3 Directive Function

The directive function entails persuading someone to do something. It refers to language that is used to cause (or prevent) overt action. This function appears

frequently in commands and requests, and used to advise, orders, suggestions, and they can be positive or negative. The directive function focuses on the addressee; it means that the speaker needs the hearer's reaction or to force someone to do something.

The aim of directive function is to give the speaker's commands. For example, "တံခါး ပိတ် လိုက် (dəgá pei? lai?, Close the door!)". To issue orders or requests, this function employs imperative statements. An imperative statement can express a strict demand, such as "တံခါး ဖွင့် လိုက် (dəgá pʰwɪ̰ lai?, open the door)", or it can appear less demanding by employing the politeness strategy, such as "တံခါး ဖွင့် ပေးပါ (dəgá pʰwɪ́ péi pà, open the door, please), or, in the case of informality, by employing question tags, such as:

Utterance_6: ဆုဆု တီဗီ ပွင့် နေ တယ်
(sʰṵ sʰṵ tì bì pwɪ̰ nèi tè, Su Su, the TV is turning on!)

3.2.4 Phatic Function

Phatic utterances express solidarity and empathy with others, for examples, "ဟေ့ နေကောင်း လား (hɛ́i nèi káʊ̃ lá , Hello, how are you?)", "ဟေ့ သူငယ်ချင်း ဘယ် လဲ (hɛ́i thange gjin: bè lé, Hey friend, where to?)". The phatic function helps in contact and refers to the communication channel. For social reasons, it either opens the channel or checks to see if it is operational. This feature is used to support social interaction.

In everyday interactions, this is one of the most common types of speech; it includes greetings, compliments, gossip, and so on. A speaker can say "မင်္ဂလာ ပါ (mɪ̰ galà pà, hello/hi)" to friends, who can be greeted with "မင်္ဂလာ ပါ (mɪ̰ galà pà, hello/hi)", but "မင်္ဂလာ မနက်ခင်း ပါ (mɪ̰ galà mane? khin: pà, Good morning)", "မင်္ဂလာ နေ့လည်ခင်း ပါ (mɪ̰ galà nɛ́i lè khin: pà, Good afternoon)", "မင်္ဂလာ ညနေခင်း ပါ (mɪ̰ galà nja. nei khin: pà, Good evening)", which are a more formal greeting.

3.2.5 Aesthetic Function

Aesthetic function refers to a focus on language's poetic features. It is the message's essence, in which the form chosen is the message's essence. Messages

convey more than just information. These additions serve no purpose other than to improve the aesthetic appeal of the messages. Poetic ability is defined as the ability to manipulate language creatively rather than writing poetry. The poetic function's goal is to convey pleasure. For example,

Utterance_7: အချိန် သည် အကောင်းဆုံး သမားတော်
(achein θè akaun: hsoun: θamá dɔ̀, Time is the best doctor)

3.3 Construction of MmTravel Corpus

The MmTravel corpus, which stands for the Myanmar Travel corpus, is derived from conversations that occur during travel. The construction of the MmTravel corpus is a contribution of this dissertation. Every utterance in the MmTravel corpus contains at least one dialogue act, indicating that it is a communicative act. For example, the conversation between speaker A and B in the corpus:

Speaker A: အိုကေ ဘယ် နှစ်စောင် လဲ (òkèi bè n̥iʔsaʊ̃ lé, Ok, how many?)

Speaker B: လက်မှတ် နှစ်စောင် ပေးပါ (leʔm̥aʔ n̥iʔsaʊ̃ péipà, Give me two tickets.)"

Which includes two utterances: "Ok, how many?" and "Give me two tickets." Each is a distinct communicative act consisting of an acknowledgement ("Ok, how many?") and a commissive ("give me two tickets "), which commits the speaker to a future action. In the traditional sense, utterances are not always grammatical sentences. Whereas a sentence is a grammatical unit composed of one or more words that follow a specific syntax, utterances, such as emoticons, are not always words; for example, ":-)" denotes a smile, which may or may not indicate acknowledgement depending on the context. There are also many single-word expressions, such as "ကောင်းပြီ (káʊ̃ pyì, ok)", "အင်း (in:, yes)", and "ဝိုး (wow)". Cue phrases are not considered utterances for the purposes of this study, and dialogue acts cannot span multiple utterances; each utterance has only one dialogue act assigned to it.

3.3.1 Corpus Creation

The MmTravel Corpus was designed to cover utterances for all potential topics in travel conversations. Since it is almost infeasible to collect them by transcribing actual conversations or simulated dialogs during the research period, utterances in our

corpus are built on the ASEAN-MT corpus, which is a collection of sentences without name entity tags that bilingual travel experts consider useful for people traveling to or returning from another country [11]. In the ASEAN-MT corpus, each of the ten different languages has a training corpus of 20K sentences, and all data sets are sentence-aligned. The sentences organized into six major categories and is also expressed in a Table 3.1.

Table 3.1 Characteristics of the ASEAN-MT Parallel Corpus

Language	Malay, Cambodian, Indonesian, Lao, Malay, Myanmar, Filipino, Chinese, Thai, Vietnamese
Size	20,000
Domain	Travel
People	Communication, Greeting, Introduction
Survival	Accommodation, Finance, Transportation
Food	Beverage, Food, Restaurant
Fun	Recreation, Shopping, Traveling
Resource	Currency, Number, Time
Special Needs	Health, Emergency

The investigation began with phrase books containing bilingual sentence pairs considered useful for tourists traveling internationally. These sentences were also collected and rewritten in a colloquial style to maximize context independence in conversation. Sentences unrelated to travel or containing highly specific meanings were excluded. Due to the utterances not originating from actual speech conversations, a broad coverage corpus could be efficiently created. The next step involved replacing names in the ASEAN-MT corpus, which includes city names, currency names, restaurant names, and other specifics related to travel in Thailand. As an example,

ASEAN-MT Corpus: ဒွန်မောင်း လေဆိပ် ကို ပါ

(don máõ lèi zeiʔ kò pà, To Donmuang Airport, please.)

MmTravel Corpus: ရန်ကုန် အပြည်ပြည်ဆိုင်ရာ လေဆိပ် ကို ပါ

(yɛ̃ kòõ ɲ pyì pyì sʰàĩ yà lèi zeiʔ kò pà,
To Yangon International Airport, please.)

ASEAN-MT Corpus: ဒီ ဘတ်စ်ကား က ဟွာဟင်း ကို သွား လား

(dì baʔská kạ̰ huahĩ̀ kò θwá lá,
Does this bus go to Huahin?)

MmTravel Corpus: ဒီ ဘတ်စ်ကား က လှည်းတန်း ကို သွား လား

(dì baʔská kạ̰ hle: tã̀ kò θwá lá,
Does this bus go to Hledan?)

The MmTravel corpus also consists of simulated dialogs between a taxi driver and a passenger, as shown below:

Driver: မင်္ဂလာ ပါ ဘာ ကူညီ ရ မလဲ

(mĩ̀ galà pà bà kù ɲì rạ̰ mạ̰ lé, Hello, what can I do for you?)

Passenger: ဘူတာရုံ ကို သွား ဖို့ ဘယ်လောက် ကြာ မလဲ

(bù dà yòʊ̃ kò θwá pʰọ̰ bè laʊʔ ʧà mạ̰ lé,
How long does it take to reach the station?)

Driver: ၄၅ မိနစ် ကြာ မယ်

(léi ze. ŋá mḭ niʔ ʧà mè, It will take 45 minutes.)

Passenger: မြန်မြန် သွား နိုင် မလား

(myã̀ myã̀ θwá nàĩ mạ̰ lá, Can you go faster?)

Driver: မြို့တွင်း က ကီလို ၄၀ ကနေ ပို့မောင်း လို့ မ ရ လို့ ပါ

(myọ̰twĩ́ kạ̰ kìlò léize kạ̰nèi pò máʊ̃ lọ̰ mạ̰ra̰ lọ̰ pà,
I must not drive over 40 km.)

Passenger: ဘယ်လောက် ပေး ရ မလဲ

(bè laʊʔ péi rạ̰ mạ̰ lé, How much?)

Driver: ၅၀၀၀ ပဲ ပေးပါ (ŋá htaun pé péi pà, 5000)

ပျော်ရွှင်ဖွယ် ခရီး ဖြစ် ပါစေ

(pyɔ̀ sw̰ĩ́ pʰwè kʰəyí pʰyiʔ pà sèi, Have a nice trip!)

The frequency distribution of corpus may differ from the “actual” frequency distribution. The frequency of an utterance in this corpus (indirectly) corresponds to the

number of travel experts who came up with the sentence and the number of situations they believe it will appear in. As a result, this frequency distribution can be thought of as a first approximation of reality. The statistics of MmTravel corpus are also seen in Table 3.2.

Table 3.2 Statistics of MmTravel Corpus

Number of Utterances	80,000
Number of Words	708,096
Number of Syllables	10,621,440
Average Utterance Length (Words)	6

3.3.2 Dialogue Act (DA) Classes

Sociolinguistic theorists initially proposed the concept of speech acts in human conversation. Dialogue Act (DA) theory suggests that in addition to conveying information, natural language utterances often express underlying intentions or actions. The first step in dialogue processing involves assigning a functional tag (DA) to user input to represent these communicative intentions. Two approaches—DA-based and Inference-Based—aim to capture these intentions, either through individual utterance analysis or holistic dialogue examination. Developing a coherent and usable taxonomy of dialogue acts remains a challenge, given varying interpretations among researchers. Standardizing theories for dialogue act identification is debated, with some advocating for precise definitions and others favoring broader frameworks of rational interaction. Recognizing dialogue acts in system-person dialogues requires robust language understanding systems, considering syntactics, semantics, and pragmatics. The term "dialogue act" has become central in computer dialogue system research and dialogue annotation.

3.3.3 Dialogue Act Mapping

Several types of dialogue systems require labeling of various kinds of acts. To create our dialogue act tag set, we manually labelled our corpus by 29 tags which are based on Myanmar speech function [89] [90] [91] [92] [93] [94] [95] is listed in Table 3.3.

When analyzing the classification of speech acts, it must be able to identify verbs used in sentences. One way to identify a specific dialogue act in DA label mapping is to use a performative verb: that is, a verb that names the speech act or illocutionary force of an utterance. The semantic base for all expressions in the set of speech act verbs is that they can be used to refer to a specific type of situation, which can be described as follows: a speaker says something to a hearer with a specific intent. Different roles are defined in the case of default: speaker role (S), hearer role (H), and the role of the utterance with its content, or, more precisely, with its propositional content (P). The set of speaker attitudes may be further specified as S's taking P to be true, S's wanting P, S's evaluating P positively or negatively, and so on. Specifications of the speaker's intention include S's intention to make H believe something or to get him/her to do something. The role of the utterance is specified by properties of the propositional content. Epistemic operator such as want and know have been used to describe the speech acts and the syntax used. The scheme with the category descriptions and examples is explained as follows.

Table 3.3 Myanmar Dialogue Act Tagsets with Speech Functions

Informational			Expressive			Directive			Phatic			Aesthetic		
1	INFORM	inf	2	ACCEPT	ac	11	COMMAND	cmd	26	GOODBYE	gb	29	AESTHETIC	as
			3	APOLOGY	apol	12	DIRECTIONS	dir	27	GREETING	gt			
			4	COMPLAIN	cp	13	INSTRUCTION	instr	28	SELF_INTRO	s_i			
			5	CONGRATULATE	cong	14	INVITE	inv						
			6	DENY	dny	15	PROHIBIT	proh						
			7	OPINION	op	16	REQUEST	req						
			8	REVIEW	rev	17	SUGGESTION	sug						
			9	THANK	thx	18	URGE	u						
			10	WISH	w	19	WARNING	wn						
						20	CHOICE QUESTION	ch_q						
						21	COMPLAIN QUESTION	cp_q						
						22	CONFIRM QUESTION	cfm_q						
						23	INQUIRY QUESTION	inq_q						
						24	OTHER QUESTION	otr_q						
						25	REQUEST QUESTION	req_q						

3.3.3.1 Informational Dialogue Act

The INFORM (inf) act is a type of informational function that is concerned with the message exchanged among the S and the H. It is a method of communicating

information through the act of stating in speech and writing. It is used to convey new information and is based on the truth and value of the information, which can be both positive and negative. It conveys information to alert someone to something, and the H can determine whether the information is correct or not. Inform is the act of S informing H of P with the precondition that S knows that P is true and the effect that hearer knows that P is true. So, the S is an authority on the subject matter of P, which can be expressed as:

know (S P)

The inform act specifies what information is to be provided in the utterance, which is a statement that the S is going to make an utterance with propositional content directed at the H. The fact that the S wants the H to inform P and hence can be represented as:

want (S know (H P))

These are some examples of inform dialogue act utterances.

Utterances_8: နှစ် နာရီ လောက် ကြာ တယ်
(n̥iʔ nà yì lɔʊʔ ʧà tè, It takes about two and a half hours)

Utterances_9: ဒီမှာ ပါ တစ် ဈောင် တစ် သောင်း ခွဲ ပါ
(dìm̥à pà tiʔ saʊ̃ tiʔ θáʊ̃ kw̥é pà, Here, 15,000 kyats per ticket)

We also mapped the inform act when the S wants to inform the H that he will do P. This gives the following function with its example utterance.

want (S know (H will (do (S P)))

Utterances_10: ကျွန်တော် ဒေါ်လာ နဲ့ ပေး မယ်
(ʧənɔ̀ dɔ̀ là nɛ̰ péi mè, I will pay in dollars)

3.3.3.2 Expressive Dialogue Act

Expressive dialogue act is an expressive speech function to expresses on the S's attitudes and emotion towards the proposition. It must be analyzed because act cannot be separated from human beings in everyday conversation. In MmTravel corpus, there are nine types of dialogue act in expressive function which are ACCEPT (ac), APOLOGY (apol), COMPLAIN (cp), CONGRATULATE (cong), DENY (dny),

OPINION (op), REVIEW (rev), THANK (thx) and WISH (w). In fact, the S wants the H to understand how he/she feels about P which can be denoted as:

want (S recognize (H feels (S because (of (P)))))

The following are example of utterances for each expressive dialogue act:

ACCEPT (ac) dialogue act refers to the act of expressing an opinion to someone or agreeing to do something or say "ဟုတ်ကဲ့ (hoʊʔ kɛ̰, yes)" to something. We assume that P is accepted with the implied sense that the S wants the H to do P. This produces the function below, along with an example utterance.

Utterances_10: ရ ပါ တယ် ပြော ပါ
(rə̰ pà tè pyɔ́ pà, Ok, go on)

The human desire to express regret for offenses is served by an apology. According to Goffman (1971), an apology is an action that attempts to uphold social norms and bring about harmony; it also helps to mend relationships. According to Searle, APOLOGY (apol) dialogue act includes the class of "expressive" speech acts. As a result, it can be used as a tool for conveying regret. As a result, the S expresses how they feel about other people, but it is unclear what they really mean. However, to impact the H, the S should present a feeling of regret, responsibility and remedy. In a difference sense, S expresses negative feelings toward a H to appease them. The assumption is that if the H is unable to fulfill the request, he or she will lose face, no matter how minor. For example, if we ask for something in a store and the staff is unable to provide it, a small apology is in order.

Utterance_11: တောင်းပန် ပါ တယ် အမ မှာ တာ ကုန် သွား ပြီ
(táʊ̃ bə̃ pà tè əmə̰ mə̰̀ tà kòʊ̃ θwá pyì,
Sorry, your order is out of stock)

It may consist of only one word to perform an apology, as in "ဆောရီး (hso: yí, Sorry)" or several words or sentences, as:

Utterance_12: နောက်ကျသွား တာ တောင်းပန် ပါ တယ်
(naʊʔ ʧə̰ θwá tà táʊ̃ bə̃ pà tè, Sorry, I'm late.)

COMPLAIN (cp) dialogue act used when the S expresses displeasure or annoyance – censure – as a reaction to a past or ongoing action P, the consequences of which S perceives as negatively affecting her. For example, when you receive poor service at a restaurant, you should make a complaint to the waiter/waitress who served you and request to speak with the manager if you are not satisfied with their response,

Utterance_13: ကြက် ပြုတ်ရည် က ရေ နဲ့ တူ တယ်
(ʧɛʔ pyoʊʔ yì kə̰ yèi nɛ̰ tù tè, Chicken broth tastes like water.)

CONGRATULATE (cong) mapped when people frequently use this speech act to express happiness or pleasure to others when something positive occurs, referring to happy events and impressive actions. The S employs the congratulatory strategy by uttering the word "ဂုဏ်ယူ ပါ တယ် (gòʊ̃ yù pà tè, Congratulations)" to H.

DENY (dny) dialogue act is a type of speech act that is delivered in response to another person's request, invitation, offer, or suggestion, implying that it is not the speaker's initiative. It is a blunt way of saying "မ ဟုတ် ဘူး (mə̰ hoʊʔ bú, No)", which is a response to an actual or implied request. It relates to an action that the H wants the S to perform P.

In everyday life, speech act is frequently necessary to distinguish between facts and opinions. Judgement as to the respective status of a particular statement is typically achieved through recourse to the "real world" where the empirical validity of the statement may be ascertained. When referring to the outside world, S and H share an evaluation procedure that can be used to determine whether a particular statement is a fact or merely an opinion. OPINION (op) dialogue act is mapped when S expresses his/her psychological state; the implementation is not an action, or particularly physical action. This is a common type of speech act that occurs in our daily lives as in the following utterances,

Utterance_14: ဒီနေ့ ရထား နောက် ကျ နေ တာ ထင် တယ်
(dìnḛ yatʰá naʊʔ ʧə̰ nèi tà tʰĩ̀ tè, I think that the train is late today)

Utterance_15: ခရီးသွား ရ တာ ဝါသနာ ပါ တယ် ။
(kʰəyí θwá yə̰ tà wà ðanà pà tè, I love to travel.)

REVIEW (rev) dialogue act is mapped by a user or consumer of a product or service based on the speaker's own experience as a user of the reviewed goods or service. This applies for reviews of food, restaurants, and hotels in our daily routine, as in the following expressions:

Utterance_16: အလွန် လတ်ဆတ် ပြီး အရသာ ရှိ သော ကမာကောင်
(əlṵ laʔ sʰaʔ pyí əyə̰ðà ʃḭ θɔ́ kəmà kaṵ,
Super fresh delicious oysters)

Utterance_17: ဈေးနှုန်း သင့်တင့် တယ်
(zéi n̥óʊ̃ θḭ tḭ tè, Reasonable price)

THANK (thx) dialogue act used when the S expresses gratitude to the person who has assisted the S, which means S feels grateful or appreciative about the action of P which has done by H. The past act action benefits S and S believes such action benefits S. This act expressions may often be required by social convention. The way gratitude is verbally expressed varies, ranging from simple, "ကျေးဇူးတင် ပါ တယ် (ʧéi zú tḭ pà tè, thank you)", or "ကျေးဇူး ပါ (ʧéi zú pà, thanks)", to the more extensive, like:

Utterance_18: အခုလို ကူညီ တာ အရမ်း ကျေးဇူးတင် ပါ တယ်
(əkʰṵ lò kù ɲì tà əyə̌ ʧéi zú tḭ̌ pà tè,
Thank you so much for your help.)

Utterance_19: မြို့ ကို လိုက် ပြ ပေး တဲ့ အတွက် ကျေးဇူးတင် ပါ တယ်
(myo̰ kò laiʔ pyə̰ péi tẹ̰ ətweʔ ʧéi zú tḭ̌ pà tè,
Thank you very much for showing me around the town.)

Every language and culture have its own set of wishes for various situations. People express their emotions about a situation, an event, or anything else. A wish is a speech act in which one expresses an opinion about another's behavior. WISH (w) dialogue act is mapped in the verbal expression of a desire or pleasure as

Utterance_20: သွား လမ်း သာ လို့ လာ လမ်း ဖြောင့် ပါစေ
(θwá lə̌ θà lo̰ là lə̌ pʰyaʊ̰̃ pà sèi, Have a safe trip!)

3.3.3.3 Directive Dialogue Act

Directive is a speech act performed by a S with the intent that the H, as well as another S in the speech or utterance, perform the action specified in the statement. For

example, “ရေ တစ် ခွက် ပေး ပါ (yèi tiʔ kʰweʔ péi pà, Please give me a cup of water)” means that the S request the H to get water. So, directive dialogue act can be mapped when the S wants to get the H to do something and hence this condition might then be:

want (S do (H P))

In MmTravel corpus, most of the dialogue act are directive dialogue act, which includes COMMAND (cmd), DIRECTIONS (dir), INSTRUCTION (instr), INVITE (inv), PROHIBIT (proh), REQUEST (req), SUGGESTION (sug), URGE (u), WARNING (wn), and six different kinds of QUESTION act, which are CHOICE QUESTION (ch_q), COMPLAIN QUESTION (cp_q), CONFIRM QUESTION (cfm_q), INQUIRY QUESTION (inq_q), OTHER QUESTION (otr_q), and REQUEST QUESTION (req_q). Here are some examples of utterances for each directive dialogue act:

A COMMAND (cmd) dialogue act is a directive that requires the H to provide information, goods, or services. A command is a method of requesting goods and services in an imperative statement, whether positive or negative. The subject is omitted in command sentences. The utterance “မြန်မြန် လာ (myɛ̃ myɛ̃ là, come faster)” is intended to command in the form of imperative statement. According to the context of the utterance, S ordered H to walk faster because S was not patient enough to walk slowly. So, S command H to walk faster. As a result, the dialogue act of this utterance is a command act.

When providing directions, the dialogue act “DIRECTIONS (dir)” is used to describe the process of guiding someone who is unfamiliar with the area to a specific location. Locator remarks are a class of descriptive utterances that serve to locate the S on a physical or mental map. To allow the H to check their position while following the directions later, a S who is delivering directions employs locator remarks. Locator remarks let the H check if they are following the directions correctly. The example utterances are as follow:

Utterance_21: ရှေ့တည့်တည့် ကို လျှောက် ပါ
(ʃḛi tḛ tḛ kò ʃaʊʔ pà, Walk straight ahead of you)

Utterance_22: ညာ ဘက် ကွေ့ ပါ
(ɲà beʔ kwḛi pà, Turn right!)

INSTRUCTION (instr) dialogue act is mapped when S uses instructing to tell H what they need to do about P. This act frequently employs imperative form which will be demonstrated in the instances below.

Utterance_23: မင်း ရဲ့ အိတ် တွေ ကို ဒီမှာ ချ ပြီး ဖွင့် လိုက် ပါ
(mɪ̀ yɛ̰ ei? twèi kò dì m̥à ʧʰa̰ pyí pʰwɪ̰̀ lai? pà,
Drop your bags here and open them.)

Utterance_24: တစ်ရာ့တစ် ကို နှိပ် လိုက်ပါ
(ti? yà ti? kò n̥ei? lai? pà, Press 101!)

By inviting, the S simultaneously commits to the suggested action and commands the H to participate in it. As a result, an INVITE (inv) dialogue act might be viewed as a directive speech act where the H takes a future action. For example, the utterance "နောက် လည်း လာလည် ပါ နော်, nɔ̰ʊ? lé làlè pà nɔ̀, please come again" where the S wants the H to come to a future event. A successful invitation is dependent on both the H accepting the offer and the S honoring the commitment made. Both the S and the H may need to engage in person for this to work.

PROHIBIT (proh) dialogue act is mapped when S prohibits or does not permit the H to do something. In the utterance "မ ထိ နဲ့ (ma̰ tʰḭ nɛ̰, Don't touch)", it has a forbidding function because it contains the illocutionary indicating device "မ - နဲ့ (Don't)". The word "don't" is an illocutionary force that indicates a prohibition act. The message delivered to the H for them to ignore the wishes of S. This utterance was intended to tell H not to do about P because S did not want H to do it.

REQUEST (req) dialogue act is a directive speech act whose illocutionary goal is to persuade the H to do something that would not be obvious in the normal course of events. The S believes that by making a request, the H can perform an action P. Possible effects of request act include firstly that the H may ask the S to do P himself. Secondly, the H may refuse or decline to do P. It is even possible that H may defer P in some way by promising to do P later, or even nominating someone else to do P. In the Utterance_25, the S is conveying a want to the H to buy a ticket, namely that the H carries out action P,

Utterance_25: လက်မှတ် တစ် စောင် ဝယ် ခဲ့ ပေး ပါ
(le? m̥a? ti? sa᷉ wè kʰɛ̰ péi pà, Please buy a ticket!)

In our corpus, the perspective of request sequences can vary as speaker-oriented and hearer-oriented. The speaker-oriented requests emphasize the role of the speaker and the hearer_oriented requests focus on the role of the hearer.

Speaker_oriented: လက်မှတ် ရ နိုင် မလား
(leʔ mạaʔ yạ̀ nàĩ mạ́ lá, Can I have a ticket?)

Hearer-oriented: မင်း လက်မှတ် ဝယ် ပေး ပါ လား
(mĩ̀ leʔ mạaʔ wè péi pà lá, Can you buy a ticket?)

The SUGGESTION (sug) falls into the category of directives, which are acts in which the S attempts to persuade the H to commit to a future course of action P. The S intends the utterance to be taken as sufficient reason for the H to do an act. When producing directive speech acts, both the S and the H must be considered. S gives the H the option of accepting or rejecting what is suggested to him/her. This means that when the S suggests to the H about P, H does not impose an obligation to accept it; rather, the H is free to accept or reject it.

Utterance_26: လားရှိုး မြို့ ကို သွား လည် သင့် တယ်
(láshou: myọ̀ kò θwá lè θĩ̀ tè, You should visit Lashio.)

URGE (u) dialogue act used when a S requests something from the H in a way that makes it clear they strongly believe it is important and are making a sincere effort to convince them to do it. These are illustrative utterances from our corpus.

Utterance_27: ခုနှစ် နာရီ ခွဲ လောက် သွား ရအောင်
(kʰṵ n̥iʔ nà yì kw̥é lɔʊʔ θwá yạ̀ aŏ̰, Let's go around 7:30.)

Utterance_28: ညစာ စောစော ပြီးအောင် စား ကြ စို့
(ɲạ̀ zà sɔ́ sɔ́ pyí aŏ̰ sá ʧạ̀ sọ̀, Let's finish dinner early.)

In the corpus, we categorized the conversation act as WARNING (wn) dialogue act when the S warns the H about the consequences and the risks that will be got if the H does that action or not. For example,

Utterance_29: နောက် တစ်ခါ စကား မ ပြော ခင် သတိထား ပါ
(nɔʊʔ tiʔ kʰà zəgá mạ̀ pyɔ́ kʰĩ̀ ðadḭ tʰá pà,
Next time, beware of before talking)

A QUESTION dialogue act is an interrogative question used to invoke confirmation or to ask a question that invites or requires a response. We can also assume a question is a type of interrogative statement that requests information from the H, which means S wants to know (find out) the answer. The S does not know the answer, i.e., does not know if the P is true, or, in case of the P function, does not know the information needed to complete the proposition truly. In Myanmar language, the question is always followed by these words: "နည်း (né)", "လော (lɔ́)", "စ (sa̰)", "လား (lá)", "လဲ (lé)", "တုန်း (tóʊ̃)", "ဦး (ù)", "လိမ့် (lḛĩ)", and "ရော (yɔ́)", at the end of the sentence. But the words "လား (lá)", and "လဲ (lé)" are the most used words in colloquial style. There are six types of question in MmTravel corpus which are CHOICE QUESTION (ch_q), COMPLAIN QUESTION (cp_q), CONFIRM QUESTION (cfm_q), INQUIRY QUESTION (inq_q), OTHER QUESTION (otr_q), and REQUEST QUESTION (req_q).

The mapping of dialogue act CHOICE QUESTION (ch_q) are those in which the answer is one of several options, which are composed of two parts that are joined by the conjunction "ဒါမှမဟုတ် (dà m̥a̰ ma̰ hoʊʔ, or)" as follow:

Utterance_30: ပြည်တွင်း လား ဒါမှမဟုတ် ပြည်ပ လား
(pyì dwĩ́ lá dàm̥a̰ma̰hoʊʔ pyì pa̰ lá, Domestic or foreign?)

However, most people without use the conjunction in daily conversation as follow example:

Utterance_31: လက်မှတ် တစ် စောင် လား နှစ် စောင် လား
(leʔ m̥aʔ tiʔ saʊ̃ lá n̥iʔ saʊ̃ lá, One or two tickets?)

In a COMPLAIN QUESTION (cp_q), the speaker expresses dissatisfaction or grievances while simultaneously seeking clarification or resolution from the hearer. This dual function embodies the complex dynamics of human communication, where the S assumes the role of both complainant and interrogator, while the H is tasked with addressing the complaint and providing relevant information or solutions. The P of the utterance serves as the focal point around which the complaint and query revolve, shaping the direction and tone of the interaction. For example,

Utterance_32: ငါ့ ပိုက်ဆံ ကို ဘာ လို့ ပြန် မ ရ တာ လဲ
(ŋa̱ paiʔ sʰa̰ kò bà lo̱ pya̰ ma̱ ya̱ tà lé,
Why can't I get my money back?)

Utterance_33: ဈေး နည်းနည်း များ မ နေ ဘူး လား
(zéi né né myá ma̱ nèi bú lá, Isn't the price a little high?)

In the context of speech acts, a CONFIRM QUESTION (cfm_q) plays a crucial role in the interaction between the S and the H, particularly concerning the P being discussed. When the S poses a confirm question, they are seeking validation or affirmation from the H regarding the accuracy or truthfulness of a specific proposition. This type of question typically arises in scenarios where clarity and mutual understanding are essential. For instance, the speaker might ask, "အစည်းအဝေး က ည၃ နာရီ လို့ ပြော နေ တာ လား (əsí əwéi ka̱ ŋa̱ nèi θóʊ̃ nà yì lo̱ pyɔ́ nèi tà lá, Are you saying that the meeting is at 3 PM?)" Here, the propositional content is the timing of the meeting, and the speaker's objective is to ensure that their understanding aligns with that of the hearer. The hearer's response, either confirming or correcting the proposition, facilitates effective communication by resolving any potential ambiguities or misunderstandings. This interaction underscores the cooperative nature of communication, where both parties actively engage to achieve a shared understanding. The following are the example utterances:

Utterance_34: ရှစ် နာရီ ဟုတ် လား
(ʃiʔ nà yì hoʊʔ lá, 8 o'clock, right?)

Utterance_35: ခင်ဗျား က ဟိုတယ် မှာ တည်း နေ တဲ့ ဧည့်သည် မ ဟုတ် လား
(kʰəmyá ka̱ hò tè m̥à té nèi te̱ ʔe̱θè ma̱ hoʊʔ lá,
You are a guest staying at the hotel, right?)

INQUIRY QUESTION (inq_q) is mapped when the function of speaking revolves around understanding how effectively a message is conveyed and interpreted. A fundamental question is how the speaker's communicative intent and the propositional content are aligned to ensure clarity and comprehension. This involves examining how S choose specific linguistic forms and structures to express their propositions, and how H decodes these forms based on their own knowledge, expectations, and the context of the interaction. The role of shared background knowledge and the dynamics of conversational context are critical in determining how

well the P is understood and acted upon by the H. Addressing these questions is essential for enhancing communication strategies and reducing misunderstandings in various communicative settings. For example;

Utterance_36: ဆယ် ဒေါ်လာ က ဘာ အတွက် လဲ
(sʰè dɔ̀ là kə̰ bà ə̰ twe? lé, What is $10 for?)

Utterance_37: ရန်ကုန် ကို ရထား လက်မှတ် ရ လား
(yə̃́ kòʊ̃ kò yatʰá le?m̥a? yə̰ lá, Can I get a train ticket to Yangon?)

The annotation of OTHER QUESTION (otr_q) is dealing with unclear questions and self-questions. Unclear questions, those that lack sufficient context or clarity, require annotators to infer the speaker's intent, which can lead to inconsistencies and potential biases in the data. For example, “ဒီ ပန်း က (dì pə̃́ kə̰, This flower)”. On the other hand, self-questions, where the speaker poses a question to themselves, often serve rhetorical purposes or reflect internal thought processes rather than seeking information from the hearer, for example, “နည်းနည်း ကျဉ်း နေ သလား လို့ (né né ɟḭ̃ nèi θà lá lo̰, Is it a little narrow?)”. Accurately annotating these self-directed questions necessitates distinguishing them from genuine information-seeking inquiries. Both scenarios demand meticulous attention to the surrounding discourse and an understanding of the speaker's objectives and context to ensure that annotations accurately reflect the intended communicative function.

Requests are attempts by the S to get the H to do something which may be very modest attempts as when S invite H to do it, or they may be very fierce attempts as when S insist that H do it. Therefore, the S makes requests to compel the H to do action in the future that will further the speaker's objectives. The REQUEST QUESTION (req_q) in which the intention of S is made explicit are those termed “conventional indirect”. By means of these formulations the S specifies his/her goal while considering the threatening nature of their request. The capacity of H and readiness to carry out the requested action. Expressions indicating capability, willingness, permission, and suggestive formulae fall under this subcategory. In the utterance, “အချိန် ကို ပြောပြ ပေး နိုင် မလား, əɟʰèĩ̀ kò pyɔ́ pyə̰ péi nàĩ mə̰ lá, Can you tell me the time?” is an example of an ability sub-strategies, which typically take the form of a question and use the modal

verbs “can”, “could”, or “may”. These are the example utterances which may found in our corpus;

Utterance_38: ၂၅ ရာခိုင်နှုန်း လျှော့ နိုင် ပါ သလား
(n̰iʔ ŋá yà gàĩ n̰óõ ʃ̰ nàĩ mn̰ lá, Could it be reduced by 25%?)

Utterance_39: ရန်ကင်း ကို ပို့ ပေး ပါ လား
(yə̃ kĩ̀ kò po̰ péi pà lá, Would you send me to Yankin?)

3.3.3.4 Phatic Dialogue Act

The sole purpose of a phatic expression is to perform a social task rather than to convey information. The phatic function is also connected to the communication channel between S and H. By establishing both a physical and psychological channel, it is intended to establish and maintain communication between them. Here are some utterance examples for each phatic dialogue act.

The GREETING (gt) speech act is one of the phatic speech acts that do not even have a propositional content that could be true or not, and they also lack sincerity conditions. Greetings are exchanged as courteous recognitions of the H who has just been encountered. When meeting or seeing a H, each S may find himself or herself expressing pleasure frequently throughout the day. As a result, ‘greeting’ is one of the most common acts in our daily lives, which is the utterance a phrase that is conventionally used to call the hearer and/or begin the communication. “ကြိုဆို ပါ တယ် (ʧ̀ozò pà tè, Welcoming)”, in which the S expresses delight at the arrival of the H. In most social situations, the question “နေကောင်း လား (nèi káõ lá, how are you?)” comes naturally, which is not intended to convey the message that S is greeting to H. Although that utterance is sometimes asked in a sincere, concerned tone that anticipates a detailed response about the respondent's current state, this must be pragmatically inferred from context and intonation.

Annotating the GOODBYE (gb) speech act involves identifying and categorizing instances where speakers signal the end of an interaction. This task requires recognizing various linguistic and non-linguistic cues that indicate a farewell, such as specific phrases: “သွား ပြီ (θwá pyì, goodbye)”, “နောက် မှ တွေ့ မယ် (nɔʊʔ m̥n̰ mə̀

twe̱i mè, see you later)", "ဂရုစိုက် နော် (gəyu̱ saiʔ nɔ̀, take care)". We must also consider contextual factors, such as the relationship between interlocutors and the setting of the conversation, which can influence the formality and type of goodbye used. Additionally, variations in cultural norms and conventions play a crucial role in how goodbyes are expressed and interpreted. By systematically annotating these elements, we can better understand the patterns and structures of closing interactions, contributing to more effective communication strategies and improved natural language processing systems that can accurately interpret and generate appropriate farewells.

A strong self-introduction is essential for making a positive first impression, as it tells the audience who you are and what you stand for. Annotating SELF_INTRO (s_i) speech acts involves systematically identifying and categorizing linguistic elements and communicative intentions in such discourse. This process focuses on identifying features like personal details, achievements, interests, and intentions conveyed by the speaker. It involves marking cues that signal the start, progression, and end of the self-introduction, and identifying speech acts that assert personal identity or social roles. This annotated data is valuable for linguistic research and provides insights into the sociocultural norms of self-presentation and interaction in various contexts. For example;

Utterance_40: ကျွန်တော် ဘိုဘို ပါ
(ɡʊ̃ dɔ̀ bò bò pà, I'm Bo Bo)

3.3.3.5 Aesthetic Dialogue Act

The aim of aesthetic communication is to capture, enchant and bind the attention of a H. The annotation of AESTHETIC (as) involves identifying and categorizing expressions that convey or evoke aesthetic experiences, feelings, or judgments. This process requires a nuanced understanding of language, as aesthetic speech acts often rely on metaphor, simile, and other figurative devices to convey sensory impressions and emotional responses. We annotate not only the explicit content but also the implicit connotations and cultural contexts that inform the aesthetic appreciation expressed in the speech act. This task is crucial for various applications, including literary analysis, sentiment analysis in creative writing, and enhancing human-computer interaction

where machines must interpret and respond to human expressions of beauty and artistic value.

3.3.4 Understanding and Labeling Ambiguous Dialogue Act

In developing the MmTravel Corpus, addressing the challenges of dialogue act ambiguity in Burmese discourse has been prioritized. Recognizing the task's complexity, rigorous annotation has been engaged in, leveraging linguistic expertise and cultural insights to understand travel-related communication nuances. The process involves careful annotation and collaborative discussions to clarify ambiguous utterances by considering contextual cues, speaker intentions, and linguistic conventions specific to Burmese. Despite the inherent subjectivity and challenges, a commitment to refining guidelines and methodologies has been maintained to ensure consistent and accurate dialogue act labeling. By addressing these ambiguities, the aim is to enhance the MmTravel Corpus's utility and reliability for linguistic research and natural language processing applications.

Ambiguous dialogue act utterances in Burmese present a unique challenge due to the language's rich phonetic and tonal characteristics. For example, the two Burmese utterances “မ လုပ် ပါ ရစေ နဲ့ (mə̰ louʔ pà yə̰ sèi nɛ̰, Don't do it!)” and “မ ယူ နဲ့ (mə̰ yù nɛ̰, Don't take it!)” represent different speech acts: denial and prohibition, respectively. The first phrase, “မ လုပ် ပါ ရစေ နဲ့”. This is a denial in the sense that it negates the action being considered, effectively rejecting or denying permission to perform that action. The second phrase, “မ ယူ နဲ့”, is a prohibition, which not only denies the action but also explicitly forbids it. While both utterances use the structure of negation with “မ (mə̰)”, the verbs “လုပ် (louʔ, do)” and “ယူ (yù, take)” target different actions. In a speech act perspective, the key difference lies in their intent and impact: denial simply rejects an action, while prohibition actively prevents it by imposing a restriction. Thus, the former is a softer refusal, whereas the latter carries a stronger directive force to stop an action from occurring.

Another example of the phrase “မ ဟုတ် ဘူး လား (mə̰ hoʊʔ bú lá, isn't it?)” can serve as either a confirmation question or a complaint question, depending on the context and intonation. When used as a confirmation question, “isn't it” seeks

agreement or affirmation from the hearer. It's often used when the speaker is fairly certain of something and is looking for the hearer to confirm it. For example: “ဒီနေ့ ရာသီဥတု အရမ်း လှ တယ် မ ဟုတ် ဘူး လား (dì nɐ̰i yà ðì u̯ du̯ əyɪ̃ l̥ɑ̰ tè ma ho' bu lar, The weather is beautiful today, isn't it?)”. Here, the speaker is seeking confirmation that the weather is indeed beautiful. When used as a complaint question, “isn't it” expresses dissatisfaction or annoyance. The speaker uses it to highlight a problem or something negative, often expecting the listener to agree with the complaint. For example: “ဒီ လုပ်ငန်း စဉ် ကြာ နေ တာ မ ဟုတ် ဘူး လား (dì loʊʔ ŋɪ̃ sɪ̃ ʤà nèi tà ma ho' bu lar, This process is taking too long, isn't it?)”. In this case, the speaker is complaining about the length of the process and expects the listener to acknowledge the issue. The key difference lies in the context and how the phrase is delivered. The same words can serve different purposes based on how they are used.

Ambiguous annotation between inquiry questions and request questions poses a significant challenge in linguistic analysis. Inquiry questions seek information or clarification and are typically marked by interrogative syntax, while request questions aim to elicit action or assistance and often contain modal verbs or polite forms. The ambiguity arises when a question can be interpreted both as an inquiry and a request, depending on context and intonation. The example of the annotation between the inquiry question “ကျွန်တော် ကူညီ ရ မလား (ʤənɔ̀ kù ɲì ɣa̰ ma̰ lá, Can I help?)” and the request question “ကျွန်တော့် ကို ကူညီ လို့ ရ မလား (ʤənɔ̰ kò kù ɲì lo̰ ɣa̰ ma̰ lá, Can you help me?)” involves discerning the distinct pragmatic functions of each utterance within a conversation. The first question represents an offer of assistance, where the speaker inquiries about the possibility of providing help to the hearer, indicating a proactive and supportive stance. In contrast, the second question is a direct request for assistance, where the speaker seeks help from the listener, highlighting a need for support.

The example of ambiguous speech acts that straddle the line between inform and request, particularly in the context of the Burmese phrases “မုန့်ဟင်းခါး စား ချင် တယ် ပြော တာ (mo̰ʊ̃ hɪ̃ gá sá ʧʰɪ̃ tè pyɔ́ tà, I said I want to eat mohinga)” and “မုန့်ဟင်းခါး စား ချင် တယ် (mo̰ʊ̃ hɪ̃ gá sá ʧʰɪ̃ tè, I want to eat mohinga)”, involves interpreting the speaker's intention. The first phrase is relaying information about the speaker's desire, informing

the listener of their craving. However, in the second phrase, the speaker directly expresses their wish to eat mohinga, implicitly asking the listener to help fulfill this desire, whether by providing the dish or facilitating its acquisition. Annotating these speech acts involves recognizing the subtle distinctions in intent and function, which is crucial for accurate communication and effective interaction in natural language processing systems.

And the context plays a crucial role in understanding their respective dialogue acts, for example, “ဂိတ် ကြေး အဆင်သင့် ထုတ် ထား (gei? ʧéi a̰sʰĩθḭ̃ tʰoʊ? tʰá, Get the gate fee ready.)” and “ဂိတ် ကြေး အဆင်သင့် ထုတ် ထား ပါ တယ် (gei? ʧéi a̰sʰĩθḭ̃ tʰoʊ? tʰá pà tè, The gate fee is ready)”. In the Burmese language, both sentences use similar structures and vocabulary, but their intended meanings differ based on context. In the first sentence, which translates to “instruction: give advice”, the context suggests a command or directive, likely in a formal or instructional setting where someone is being told what to do. Conversely, the second sentence translates to “inform: giving information”, and the addition of “ပါ တယ် (pà tè)” adds politeness and completion to the act of informing, indicating a polite form of conveying information. The context surrounding these utterances — such as the relationship between speaker and listener, the setting, and the preceding discourse — fundamentally shapes their interpretation and the appropriate response. Context is crucial in dialogue acts because it helps determine the intention behind the speech act, guiding the appropriate linguistic and social response. Understanding context allows speakers to navigate meaning and social norms effectively in communication.

Therefore, understanding and correctly interpreting these subtle phonetic cues are crucial for effective communication in Burmese. This complexity highlights the importance of contextual knowledge and the need for advanced natural language processing tools tailored to handle the specific linguistic features of Burmese.

3.3.5 Statistical Analysis of Dialogue Act

The statistical analysis of the MmTravel corpus involves examining the frequency and distribution of various linguistic features within a dataset centered on travel-related dialogues in Myanmar (Burmese). This analysis includes quantifying occurrences of specific dialogue acts, such as requests for information, expressions of

preference, and navigational instructions. This quantitative insight helps in understanding how travelers communicate in Burmese, aiding in the development of more effective natural language processing tools tailored to the travel domain. Additionally, the analysis can reveal cultural nuances in communication styles, enhancing the design of user interfaces and automated systems for tourism and customer service applications.

The Figure 3 displays a chart of the most frequent words in a dataset. The horizontal axis represents the count of occurrences for each word, while the vertical axis lists the words themselves. At the top, the word "သွား (θwá, go)" dominates with 21,430 occurrences, followed by "မြို့ (myo̰, city)" with 14,270 occurrences. Other frequently appearing words include "ပေး (péi, give)", "လက်မှတ် (leʔ m̥aʔ, ticket)", and "နိုင်ငံ (nàĩ ŋ̃á, country)" with counts of 9,091, 6,086, and 5,125, respectively. This distribution suggests a strong emphasis on travel experiences, temporal aspects, and positive descriptors within the dataset. The significant disparity between the most frequent words and others indicates that certain key terms are central to the discourse in the "MmTravel" dataset. This analysis highlights the focus areas and potential thematic elements prevalent in the text data, useful for further exploration and understanding of travel-related content.

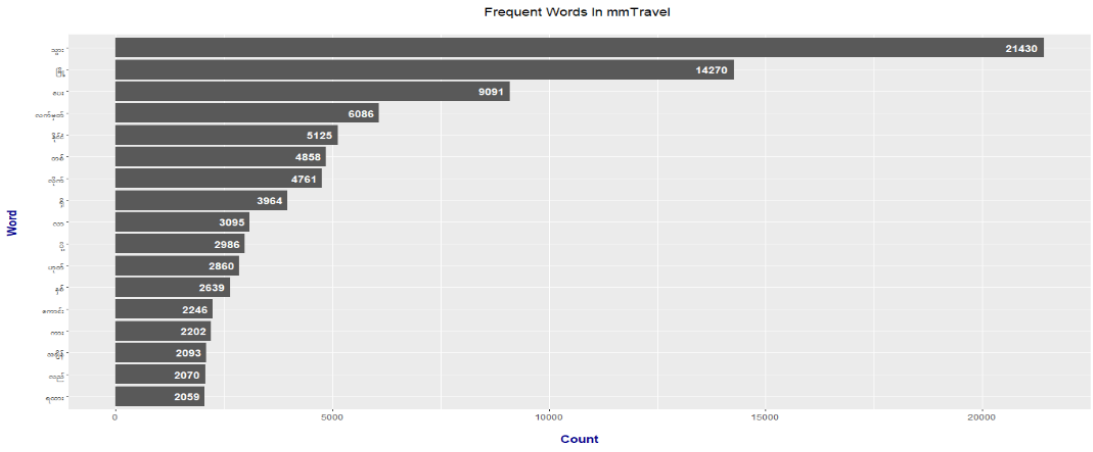


Figure 3.3 Frequencies of occurrence words in MmTravel corpus

The analysis of dialogue act frequency distribution in the MmTravel corpus is shown in Table 3.4, which reveals significant insights into the communicative dynamics of travel-related interactions. The data on Dialogue Act (DA) frequencies reveal distinct patterns in communication, emphasizing the prevalence of certain types of interactions

over others. Inquiry questions, with a frequency of 20.21%, dominate the dataset, indicating that seeking information is the most common communicative act. This is closely followed by the act of informing, which accounts for 18.83% of the instances, suggesting that providing information is almost as crucial as requesting it. Self-introduction, making up 17.33%, also ranks highly, highlighting the importance of establishing identity in conversations.

Table 3.4 Dialogue Acts (DA) Frequency Distribution in MmTravel Corpus

DA	Frequency (%)	DA	Frequency (%)
inquiry_quesiton	20.21	invite	0.99
inform	18.83	other_question	0.72
self_intro	17.33	apology	0.69
request	6.43	review	0.65
congratulate	5.74	aesthetic	0.62
opinion	3.84	complain	0.52
accept	3.65	thank	0.52
deny	3.32	greeting	0.47
suggestion	3.26	directions	0.41
request_question	2.61	choice_question	0.35
command	2.01	wish	0.27
instruction	1.88	goodbye	0.14
urge	1.57	complain_question	<0.1
prohibit	1.37	warning	<0.1
confirm_question	1.34		

Other acts like requests and congratulations show moderate frequencies at 6.43% and 5.74% respectively, reflecting common social interactions where assistance or acknowledgment of achievements are involved. Opinion sharing, acceptance, and denial, each occurring between 3.32% and 3.84%, illustrate a moderate presence of evaluative and decision-making elements in dialogues. Suggestions and request questions, falling around 3.26% and 2.61%, suggest a need for input and further information gathering, albeit less frequently than straightforward inquiries or information sharing.

Less frequent acts include commands (2.01%), instructions (1.88%), and urgings (1.57%), indicating that directive speech acts are not as dominant. Prohibitions, confirm questions, and invitations range from 1.37% to 0.99%, suggesting these specific communicative needs arise but are less common. Rare acts such as apologies,

reviews, and aesthetic comments (each under 0.70%) reflect the specialized nature of these interactions.

The data also highlight extremely rare dialogue acts like greetings, directions, choice questions, wishes, and goodbyes, each occurring less than 0.50% of the time. This rarity underscores their specific use cases compared to more frequent conversational acts. The least common acts, including complain questions and warnings, occurring less than 0.1%, are likely only used in very particular contexts, indicating their specialized and infrequent nature in regular communication. Overall, the distribution showcases a rich of information-seeking and providing acts, consistent with the nature of travel-related conversations, while also highlighting the variety of communicative actions users employ to achieve their conversational goals.

3.4 Summary

This chapter provides a comprehensive overview of the creation and analysis of a dialogue corpus specific to the Myanmar language, which is an essential resource for language processing. It begins with a general background on the Myanmar language, covering its linguistic features and the cultural context that influences communication patterns. Then delves into Myanmar speech functions, categorizing them into informational, expressive, directive, phatic, and aesthetic functions, illustrating the diverse communicative purposes in Myanmar conversations. The core of the chapter focuses on the MmTravel Corpus, detailing the meticulous process of its creation. It outlines the collection, annotation, and structuring of the data to ensure a representative and functional corpus. The dialogue act classes are introduced, describing the 29 distinct dialogue acts identified within the corpus.

A significant section is dedicated to dialogue act mapping, explaining how these acts are systematically categorized and the challenges of labeling ambiguous dialogue acts. The chapter highlights strategies for understanding and accurately labeling such ambiguities, ensuring the corpus's reliability and utility. Lastly, a statistical analysis of dialogue acts is presented, offering insights into the frequency distribution and patterns of use, which are crucial for both linguistic studies and practical applications in natural language processing. This thorough exploration underscores the chapter's contribution to understanding and leveraging the complexities of Myanmar language dialogues in the context of travel-related interactions.

CHAPTER 4

STUDY ON HEURISTIC AND MACHINE LEARNING BASED MDAR

The implementation of machine learning models before exploring into deep neural models for dialogue act classification in the Myanmar language is deemed crucial for several reasons. Firstly, machine learning models are generally less complex and fewer computational resources are required, making them more accessible for initial experimentation and understanding of the data. A strong baseline performance can be provided by them, aiding in the identification of key features and patterns within the language data, which is particularly important for under-resourced languages like Myanmar. Additionally, greater ease of interpretation is offered by machine learning models, allowing insights to be gained into the factors influencing classification results. This interpretability is essential in the refinement of the dataset and feature selection, ensuring that the data is well-prepared for more advanced techniques. By starting with machine learning models, an iterative improvement of the approach can be achieved, building a solid foundation that ultimately leads to more effective and robust deep neural models for dialogue act classification. In this chapter, two machine learning approaches were expressed: SVM with different kernels and the Naive Bayes classifier.

4.1 Naïve Bayes (NB) Classifier

Naïve Bayes classifiers [5] are a family of probabilistic models based on Bayes' theorem, which assume independence between the features given the class label. This simplification makes them particularly effective for high-dimensional datasets. In the context of dialogue act classification, a Naïve Bayes classifier can be employed to predict the type of a dialogue act based on features extracted from the dialogue text. Bayes' theorem forms the foundation of the Naïve Bayes classifier and is given by:

$$P(C_k \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid C_k) \cdot P(C_k)}{P(\mathbf{X})} \tag{4.1}$$

where: $P(C_k \mid \mathbf{X})$ is the posterior probability of class C_k given the feature vector $\mathbf{X}$, $P(\mathbf{X} \mid C_k)$ is the likelihood of the feature vector $\mathbf{X}$ given class C_k, $P(C_k)$ is the prior probability of class C_k , $P(\mathbf{X})$ is the marginal likelihood of the feature vector $\mathbf{X}$.

Among the various types of Naïve Bayes classifiers, such as MultinomialNB and BernoulliNB, are robust tools for dialogue act classification in natural language processing. These algorithms leverage Bayes' theorem and the naïve assumption of feature independence to predict the dialogue act type based on textual features extracted from conversations.

4.1.1 Multinomial Naïve Bayes (MultinomialNB)

MultinomialNB is well-suited for text classification tasks where features are typically word frequencies or term frequencies. In the context of dialogue act classification, each dialogue utterance can be represented as a bag-of-words model, where the feature vector **X** consists of word counts. The probability of observing a feature vector **X** given class C_k is modeled as:

$$P(\mathbf{X} \mid C_k) = \prod_{i=1}^{n} P(X_i \mid C_k)^{X_i} \tag{4.2}$$

Here, X_i represents the count of word i in the dialogue utterance, and $P(X_i \mid C_k)$ is the probability of word i appearing in dialogue acts of class C_k. This probability is estimated using:

$$P(X_i \mid C_k) = \frac{N_{ik} + \alpha}{N_k + \alpha n} \tag{4.3}$$

where N_{ik} is the number of times word i appears in dialogue acts of class C_k, N_k is the total number of words in all dialogue acts of class C_k, n is the total number of unique words in the vocabulary, α is the smoothing parameter (often Laplace smoothing) to handle unseen words. The prior probability $P(C_k)$ is estimated as:

$$P(C_k) = \frac{N_k}{N} \tag{4.4}$$

where N_k is the number of dialogue acts of class C_k and N is the total number of dialogue acts. During classification, MultinomialNB selects the class C_k that maximizes the posterior probability:

$$\hat{C} = \arg\max_{C_k} (\log P(C_k) + \sum_{i=1}^{n} X_i \log P(X_i \mid C_k)) \tag{4.5}$$

4.1.2 Bernoulli Naïve Bayes (BernoulliNB)

In contrast to MultinomialNB, BernoulliNB is suitable when features are binary indicators, typically representing the presence or absence of words in dialogue acts. For dialogue act classification, this model assumes a binary feature vector **X**, where X_i equals 1 if word i is present in the dialogue act and 0 otherwise. The likelihood of observing a feature vector **X** given class C_k in BernoulliNB is:

$$P(\mathbf{X} \mid C_k) = \prod_{i=1}^{n} P(X_i \mid C_k)^{X_i} \cdot (1 - P(X_i \mid C_k))^{(1-X_i)} \tag{4.6}$$

Here, $P(X_i \mid C_k)$ represents the probability that word i is present in dialogue acts of class C_k, estimated as:

$$P(X_i \mid C_k) = \frac{N_{ik} + \alpha}{N_k + 2\alpha} \tag{4.7}$$

where N_{ik} is the number of dialogue acts of class C_k containing word i, N_k is the total number of dialogue acts of class C_k, and α is the smoothing parameter. The prior probability $P(C_k)$ is similarly estimated as in MultinomialNB:

$$P(C_k) = \frac{N_k}{N} \tag{4.8}$$

During classification, BernoulliNB selects the class C_k that maximizes the posterior probability:

$$\begin{aligned} \hat{C} = \arg\max_{C_k} (\log P(C_k) + \sum_{i=1}^{n} X_i \log P(X_i \mid C_k) \\ + \sum_{i=1}^{n} (1 - X_i) \log (1 - P(X_i \mid C_k))) \end{aligned} \tag{4.9}$$

MultinomialNB and BernoulliNB are effective choices for dialogue act classification due to their ability to handle different types of textual features. MultinomialNB excels with term frequencies while BernoulliNB is appropriate for binary features indicating word presence. Their training algorithms, based on Bayes' theorem and the naïve independence assumption, provide efficient and interpretable models for predicting dialogue act types from conversational data. These classifiers are widely applied in natural language processing tasks where robust performance and computational efficiency are critical.

4.2 Support Vector Machine (SVM) Classifier

Support Vector Machines (SVM) [14] are powerful supervised learning models used for classification and regression tasks. For Dialogue Act Classification, SVMs can be employed to categorize segments of conversation into predefined dialogue acts such as statements, questions, commands, etc. The performance of SVM models largely depends on the choice of the kernel, which defines the transformation of input data into higher-dimensional space to make it linearly separable. Here, we discuss three commonly used kernels: Linear, Radial Basis Function (RBF), and Polynomial (Poly) kernels.

4.2.1 SVM with Linear Kernel

The linear kernel is the simplest type of kernel, used when the data is linearly separable. The decision function of an SVM with a linear kernel is defined as:

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b \tag{4.10}$$

where $\mathbf{w}$ is the weight vector and b is the bias. For Dialogue Act Classification, this means that the model finds a hyperplane that maximizes the margin between different dialogue acts. The optimization problem can be expressed as:

$$\min_{\mathbf{w},b} \frac{1}{2} \| \mathbf{w} \|^2 \tag{4.11}$$

subject to the constraints $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i$ for $i = 1, \ldots, n$ where $\xi_i \geq 0$ are slack variables that allow some misclassifications. The linear kernel is effective when the feature space is already well-structured, but it might not perform well for more complex distributions of dialogue acts.

4.2.2 SVM with RBF Kernel

The Radial Basis Function (RBF) kernel, also known as the Gaussian kernel, is widely used for its ability to handle non-linear relationships. The RBF kernel is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2\right) \tag{4.12}$$

where γ is a parameter that controls the width of the Gaussian function. The decision function becomes:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i \exp\left(-\gamma \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2\right) + b \tag{4.13}$$

For Dialogue Act Classification, the RBF kernel allows the model to create complex boundaries between different classes, effectively capturing the nuances in conversational data. The hyperparameter γ must be carefully tuned to balance the trade-off between bias and variance.

4.2.3 SVM with Polynomial Kernel

The Polynomial kernel can model interactions up to a specified degree d. It is defined as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + c)^d \tag{4.14}$$

where c is a constant that controls the influence of higher-order versus lower-order terms, and d is the degree of the polynomial. The decision function using a polynomial kernel is:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}_j + c)^d + b \tag{4.15}$$

In the context of Dialogue Act Classification, the polynomial kernel can capture complex patterns in the data, especially when the relationship between features and classes is more intricate. The choice of d and c needs to be optimized to avoid overfitting or underfitting.

SVMs, with their ability to utilize different kernels, provide a flexible framework for Dialogue Act Classification. The choice of kernel—linear, RBF, or polynomial—depends on the nature of the dialogue data and the complexity of the relationships between different dialogue acts. By transforming the input space in various ways, SVMs can effectively separate classes and achieve high classification accuracy, making them a valuable tool in the field of natural language processing.

4.3 Legacy Word Representation

This section reviews common legacy word representation methods that traditionally rely on word frequency. It first discusses categorical word representation methods, including one-hot encoding and bag-of-words (BoW). One-hot encoding represents each word as a binary vector with a single 1 and the rest 0s, corresponding to the word's index in the vocabulary. Bag-of-words extends this by summing the one-

hot vectors of all words in a sentence to capture word frequency. The section then moves on to weighted word representation techniques.

The two widely used weighted word representation methods are Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF). Unlike categorical representations, these methods use numerical values based on word frequency. Term Frequency (TF) measures how often a word appears in a document, adjusting for the document's length to ensure comparability. TF is calculated by dividing a word's count by the total number of words in the document. To address the dominance of common words, Sparck Jones et al. [50] proposed the inverse document frequency (IDF), which reduces the weight of frequent words and increases the weight of rarer ones. TF-IDF combines these principles to differentiate important words from common ones, enhancing text representation quality.

In theoretical perspective, the underlying structure of the text is hierarchical. Characters combine into words, words combine into sentences, and sentences combine to form documents. Documents can then be organized into collections. Different NLP tasks investigate problems at various levels of granularity; in our research, we are primarily interested in word-level and sentence-level representation. The letters t, d, and D to represent a word (term), a sentence, and a document, respectively. Given a vocabulary V (a finite list of indexed words), a word t can be represented as a one-hot encoded vector x_t of size $|V|$, with the element corresponding to the vocabulary index of t being one and all the other elements being zero. A sentence d is then represented as a weighted sum of the vectors for all the words in the sentence, as shown below:

$$x_d = \sum_{t \in d, d \epsilon D} f(t, d, D)\, x_t \tag{4.16}$$

where $f(t, d, D)$ is a weighting function for the word t in its context of d and D. The most basic approach is to set $f(t, d, D)$ equal to the number of occurrences of t in d, which is known as the Term Frequency weighting function $TF\ (t, d)$ [32]. Thus, the significance (weight) of t in d is simply proportional to its frequency. The Bag-of-Words (BoW) model is a text representation technique that represents text as a bag of words, that is, an unordered set of words that does not keep track of where a word appeared in a sentence or document D [15]. Despite the information loss caused by the

term independence assumption, which states that the presence of one word in the bag is unrelated to the presence of another, BoW can be surprisingly effective on some tasks and is widely used as a simple baseline.

The inability of BoW to capture syntactic and semantic relationships between words is its primary limitation. Despite being semantically and syntactically identical, the words "အမေ (əmèi, mom)" and "မိခင် (mḭ gḭ̃, mother)" are distinct in the view of BoW (e.g., the cosine similarity between their vectors is zero). Furthermore, the sentence "နေပြည်တော် က မြန်မာ ရဲ့ မြို့တော် (nèipyìtɔ̀ kə̰ myamà yḛ̀ myo̰dɔ́, Naypyitaw is the capital of Myanmar.)" is treated as identical to "မြန်မာ ရဲ့ မြို့တော် က နေပြည်တော် (myamà yḛ̀ myo̰dɔ́ kə̰ nèipyìtɔ̀, The capital of Myanmar is Nay Pyi Taw.)" because BoW ignores word order, which is critical in determining sentential meaning. Furthermore, output vectors are inherently high-dimensional and sparse due to the large vocabulary size, which can lead to the dimensionality curse for machine learning models [67].

This next section concentrates on the most widely used TF-IDF weighting function for the BoW model, which generates orthogonal vectors for all words: "အမေ (əmèi, mom)" vs. "မိခင် (mḭ gḭ̃, mother)". Other weighting functions developed in the literature for distributional semantic models, such as the Positive Pointwise Mutual Information (PPMI) weighting function for the BoW model, can create similar vectors for similar words (i.e., capturing lexical semantic similarity) [55].

4.4 Workflow of Machine Learning based MDAR

As detailed in Chapter 3, the MmTravel corpus served as the foundational dataset for modeling dialogue act recognition in the Myanmar language. The overview work flow of machine learning based MDAR modeling and the implementation of the system is illustrated in Figure 4.1.

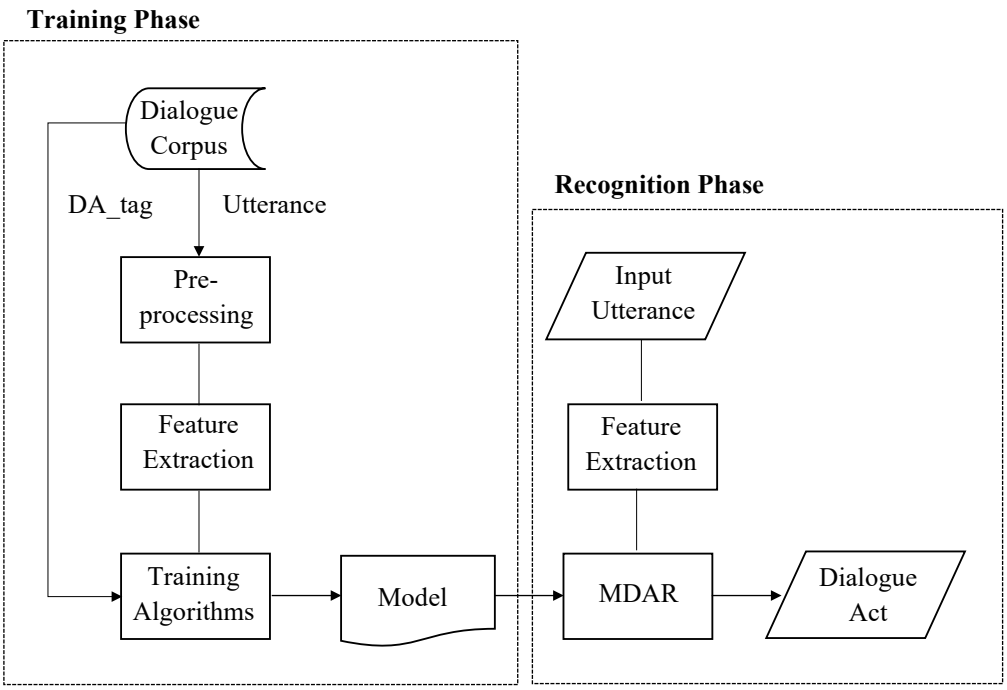


Figure 4.1 Flow chart of the proposed Myanmar Dialogue Act Recognition

4.4.1 Preprocessing and Feature Extraction

Pre-processing is a crucial step in developing machine learning models, especially for tasks like dialogue act recognition. For Myanmar dialogue act recognition using Naive Bayes and Support Vector Machine (SVM) algorithms, pre-processing ensures the text data is clean, standardized, and ready for feature extraction and model training. Given the complexity and unique characteristics of the Myanmar language, this process involves several specialized steps. The first step typically involves text normalization, which includes converting all characters to a consistent case, removing punctuation, and handling various linguistic nuances. Following normalization, tokenization is performed using a Myanmar word segmentation tool, specifically the one developed by the UCSY NLP Lab. This tool effectively segments continuous Myanmar text into meaningful words, which is essential because Myanmar script does not use spaces between words, making segmentation a non-trivial task.

Once the text is segmented into words, additional pre-processing steps are applied to prepare the data for Naive Bayes and SVM algorithms. These steps include removing stop words, which are common words that do not contribute significantly to the meaning, and stemming or lemmatization, which reduces words to their base or root form. Feature extraction follows, converting the textual data into numerical vectors that

machine learning algorithms can process. For Naive Bayes, features are often represented as term frequency-inverse document frequency (TF-IDF) vectors, capturing the importance of words within the documents. For SVM, similar feature vectors can be used, or other techniques like word embeddings might be employed to capture semantic meanings.

After pre-processing and feature extraction, the dataset is split to evaluate the performance of the models. The dataset is divided into three parts: 80% for the training set, 10% for the test set, and 10% for the validation set. The training set is used to train the Naive Bayes and SVM models, where the algorithms learn the patterns and relationships in the data. The test set is employed to evaluate the models' performance, providing an unbiased assessment of how well the models generalize to unseen data. The validation set, on the other hand, is used during the training process to fine-tune model parameters and prevent overfitting, ensuring that the models perform optimally on new, unseen data.

4.4.2 Training

A thorough analysis was conducted on MultinomialNB and BernoulliNB for the Naïve Bayes classifier, alongside various kernel-based Support Vector Machine (SVM) models on the annotated data. The process of model training in our experiment involved a meticulous examination of various hyperparameters to optimize performance. Initially, the data was loaded, preprocessed, and then split into training, testing and validation sets with an 80-10-10 ratio. Text representation was handled using the TfidfVectorizer to capture the nuanced importance of terms within the dataset. Different hyperparameters were explored for each model, focusing on Naive Bayes classifiers (MultinomialNB and BernoulliNB) and various kernel-based Support Vector Machine (SVM) models (linear, RBF, and polynomial kernels).

For the Naive Bayes models, key hyperparameters included the smoothing parameter alpha, set to 0.01 to handle the issue of zero probabilities, and the fit_prior option, ensuring the algorithm considers the class distribution in the training data. For the SVM models, kernel functions—linear, RBF, and polynomial—were experimented with to determine which best mapped the data into a higher-dimensional space for better separability. The RBF kernel, with its gamma parameter set to 'scale', demonstrated the strongest learning ability, effectively capturing complex relationships in the data.

The models were rigorously evaluated using the cross-validation method, ensuring robust performance across different subsets of the data. By systematically varying these hyperparameters and employing cross-validation, the optimal configurations that yielded the best classification results were identified. This comprehensive approach ensured that the models were not only accurate but also generalizable to new, unseen data. The results highlighted the importance of carefully tuning hyperparameters to achieve high performance in dialogue act recognition tasks.

Some of the top features extracted by the MultinomialNB and BernoulliNB classifiers on our datasets are listed in Tables 4.1 and 4.2, respectively. Among these features, common words in the Myanmar language such as "ကျွန်တော် (I)", "မင်း (You)", and "ခင်ဗျား (You)" frequently appear in the corpus. These frequent occurrences suggest the prominence of these words in various dialogue acts.

Table 4.1 Top Features of MultinomialNB Naïve Bayes

Dialogue Act		**Features Words**
Request Question	req_q	ဝမ်းနည်း ၊ နိုင် ၊ လုပ် ၊ ပေး
Directions	dir	ဒီ ၊ လမ်း ၊ ဒီမှာပဲ ၊ ဆက် ၊ သွား ၊ နား ၊ နေရာ ၊ ဘက်
Command	cmd	ထည့် ၊ ဖို့ ၊ ရဘူး ၊ နော် ၊ တော့ ၊ ကို ၊ မ ၊ နဲ့
Greeting	gt	ဟဲလို ၊ မင်္ဂလာ ၊ **ကျွန်တော်** ၊ ဟိုင်း ၊ ဝမ်းသာ ၊ ရဲ့ ၊ နေကောင်း ၊ **ခင်ဗျား**

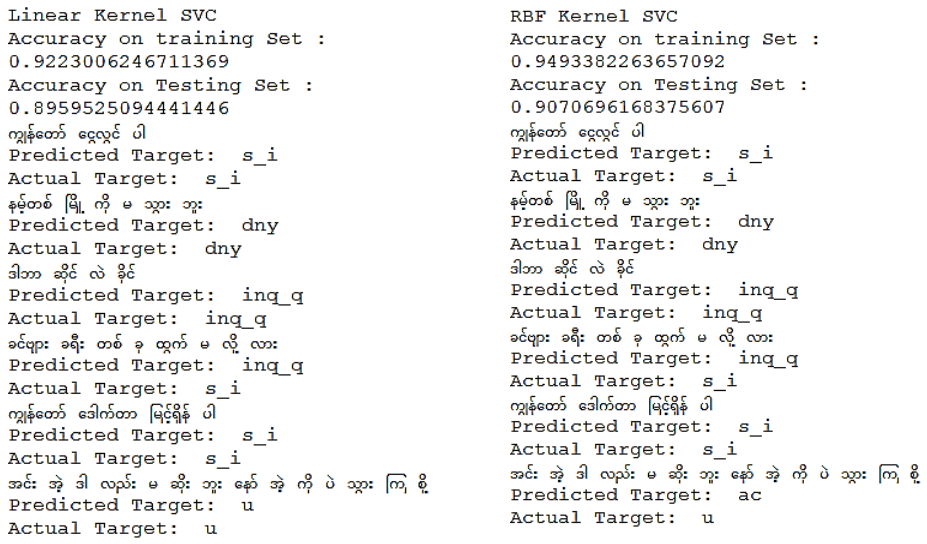
Table 4.2 Top Features of BernoulliNB Naïve Bayes

Dialogue Act		**Features Words**
Instruction	instr	သောက် ၊ ပါ ၊ **မင်း** ၊ က ၊ ဖို့ ၊ ဒီ ၊ နဲ့ ၊ ကို
Apology	apol	အားနာ ၊ စိတ်မကောင်း ၊ တောင်းပန်
Deny	dny	မဟုတ် ၊ ဘူး ၊ မဟုတ် ၊ မ ၊ ဟင့်အင်း ၊ နဲ့
Wish	w	မျှော်လင့် ၊ ပျော်ရွှင် ၊ မှာ ၊ ခရီး ၊ **ကျွန်တော်** ၊ ပါစေ

Given the frequent appearance of these common words in our corpus, our future work will involve estimating classification performance by removing stopwords from the utterances. This approach aims to refine the classification process by eliminating commonly used words that may not contribute significantly to the semantic meaning of dialogue acts.

4.4.3 Experimental Result of Support Vector Machine

Support Vector Machine (SVM) is a robust supervised learning method known for its effectiveness in classification problems, especially with linearly inseparable data. SVM achieves high accuracy by mapping data into a high-dimensional space using kernel functions. In the experiment, we examined the impact of different kernel functions were examined: the linear kernel, Radial Basis Function (RBF) kernel, and polynomial kernel. Each kernel has unique advantages: the linear kernel is efficient for linearly separable data, the polynomial kernel captures complex relationships, and the RBF kernel excels at handling non-linear data by mapping it into an infinite-dimensional space. The results showed that the RBF kernel demonstrated the strongest learning ability, consistently outperforming the linear and polynomial kernels in terms of accuracy and generalization. This superior performance underscores the importance of selecting appropriate kernel functions in SVM to enhance model effectiveness in complex classification tasks.



(a) (b)

Figure 4.2 Comparison Result of the training and testing between (a) SVM linear kernel and (b) SVM RBF kernel

Figure 4.2 shows the comparison between the SVM with a linear kernel (left side) and the SVM with an RBF kernel (right side) shows differences in performance and accuracy on both training and testing sets. The linear kernel SVM achieved an accuracy of 92.23% on the training set and 89.59% on the testing set, whereas the RBF kernel SVM demonstrated higher accuracy, reaching 94.93% on the training set and 90.77% on the testing set. This indicates that the RBF kernel SVM has a better

generalization ability compared to the linear kernel SVM. Additionally, examining the predicted and actual targets, the RBF kernel SVM appears to have fewer misclassifications, which further corroborates its superior performance. The results suggest that for this particular classification task, the RBF kernel is more effective in capturing the underlying patterns in the data, leading to better overall accuracy.

The comparison between the SVM with a linear kernel (left side) and the SVM with an RBF kernel (right side) for each dialogue act demonstrates varying performance across different metrics such as precision, recall, and F1-score, is shown in Figure 4.3. The overall accuracy of the RBF kernel (91%) is slightly higher than the linear kernel (90%). For dialogue acts such as "ac," "apol," "cfm_q," and "cmd," the RBF kernel shows better precision, recall, and F1-score, indicating more reliable predictions.

Classification Report :

	precision	recall	f1-score	support
ac	0.85	0.76	0.80	673
apol	0.89	0.87	0.88	126
as	0.71	0.49	0.58	107
cfm_q	0.95	0.90	0.93	235
ch_q	0.62	0.25	0.36	63
cmd	0.96	0.93	0.95	352
cong	1.00	1.00	1.00	1069
cp	0.19	0.05	0.08	81
cp_q	0.60	0.17	0.26	18
dir	0.66	0.39	0.49	80
dny	0.84	0.83	0.83	636
gb	0.70	0.39	0.50	18
gt	0.85	0.75	0.79	83
inf	0.78	0.90	0.83	3535
inq_q	0.94	0.97	0.96	3812
instr	0.88	0.70	0.78	376
inv	0.98	0.95	0.96	182
op	0.78	0.54	0.64	721
otr_q	0.69	0.53	0.60	126
proh	0.88	0.92	0.90	267
req	0.91	0.91	0.91	1158
req_q	0.91	0.86	0.89	462
rev	0.58	0.38	0.46	118
s_i	0.99	1.00	1.00	3173
sug	0.96	0.91	0.93	599
thx	0.93	0.98	0.96	100
u	0.97	0.97	0.97	291
w	0.88	0.93	0.90	55
wn	0.80	0.29	0.42	14
accuracy			0.90	18530
macro avg	0.82	0.71	0.74	18530
weighted avg	0.89	0.90	0.89	18530

(a)

Classification Report :

	precision	recall	f1-score	support
ac	0.87	0.80	0.84	673
apol	0.93	0.87	0.90	126
as	0.89	0.31	0.46	107
cfm_q	0.95	0.92	0.94	235
ch_q	0.73	0.30	0.43	63
cmd	1.00	0.92	0.96	352
cong	1.00	1.00	1.00	1069
cp	0.31	0.06	0.10	81
cp_q	1.00	0.11	0.20	18
dir	0.81	0.36	0.50	80
dny	0.87	0.85	0.86	636
gb	0.80	0.44	0.57	18
gt	0.87	0.71	0.78	83
inf	0.77	0.93	0.84	3535
inq_q	0.95	0.98	0.97	3812
instr	0.93	0.68	0.79	376
inv	0.99	0.95	0.97	182
op	0.83	0.56	0.67	721
otr_q	0.77	0.57	0.65	126
proh	0.92	0.93	0.92	267
req	0.92	0.93	0.92	1158
req_q	0.97	0.90	0.93	462
rev	0.71	0.35	0.47	118
s_i	1.00	1.00	1.00	3173
sug	0.98	0.90	0.94	599
thx	0.92	0.98	0.95	100
u	0.98	0.97	0.97	291
w	0.83	0.91	0.87	55
wn	1.00	0.07	0.13	14
accuracy			0.91	18530
macro avg	0.88	0.70	0.74	18530
weighted avg	0.91	0.91	0.90	18530

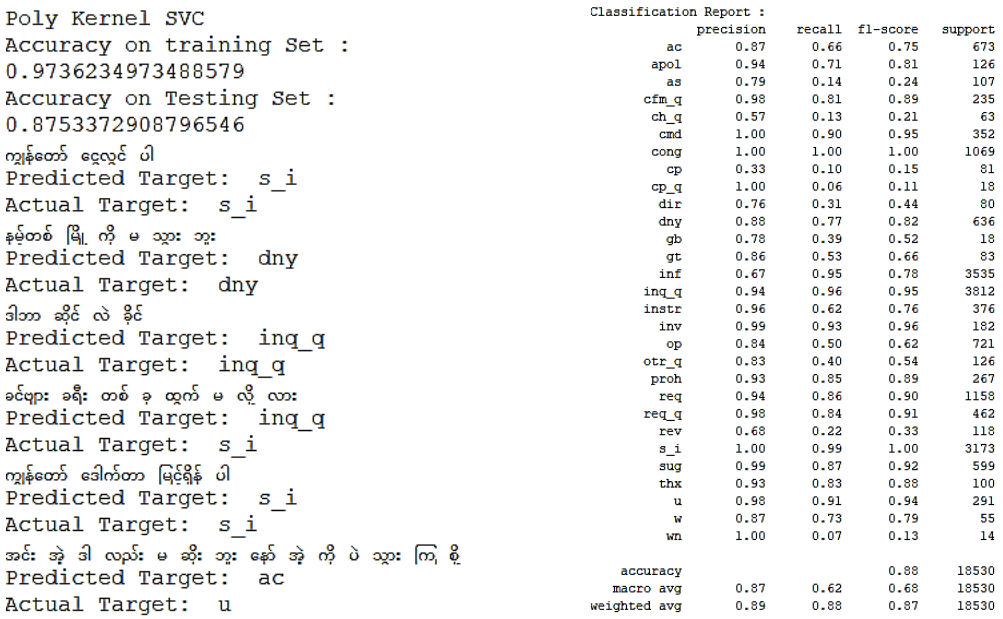
(b)

Figure 4.3 Comparison Result for each dialogue act between (a) SVM linear kernel and (b) SVM RBF kernel

Specifically, for the dialogue act "ac," the precision improved from 0.85 to 0.87, recall from 0.76 to 0.80, and F1-score from 0.80 to 0.84 with the RBF kernel. Similarly, for "cfm_q," the precision increased from 0.71 to 0.83, recall from 0.91 to 0.92, and F1-score from 0.80 to 0.87. For dialogue acts like "s_i" and "u," both models performed very well, but the RBF kernel still showed marginal improvements.

The linear kernel SVM did perform better in a few instances, such as for the dialogue act "cong," where both precision and recall were perfect (1.00), while the RBF kernel also achieved a perfect score, making no difference in this case. However, for less frequent dialogue acts such as "cp_q" and "dny," the RBF kernel significantly outperformed the linear kernel, as evidenced by higher F1-scores.

The macro average and weighted average for the RBF kernel (0.74 and 0.91 respectively) also outperform those of the linear kernel (0.71 and 0.90 respectively), reflecting its overall superior performance across all dialogue acts. This detailed comparison suggests that the RBF kernel SVM is generally more effective and reliable for classifying dialogue acts.



```
Poly Kernel SVC
Accuracy on training Set :
0.9736234973488579
Accuracy on Testing Set :
0.8753372908796546
ကျွန်တော် ငွေလွှဲ ပါ
Predicted Target:  s_i
Actual Target:  s_i
နှမ်တစ် မြို့ ကို မ သွား ဘူး
Predicted Target:  dny
Actual Target:  dny
ဒါဘာ ဆိုင် လဲ ခိုင်
Predicted Target:  inq_q
Actual Target:  inq_q
ခင်ဗျား ခရီး တစ် ခု ထွက် မ လို့ လား
Predicted Target:  inq_q
Actual Target:  s_i
ကျွန်တော် ဒေါက်တာ မြင့်ရှိန် ပါ
Predicted Target:  s_i
Actual Target:  s_i
အင်း အဲ့ ဒါ လည်း မ ဆိုး ဘူး နော် အဲ့ ကို ပဲ သွား ကြ စို့
Predicted Target:  ac
Actual Target:  u
```

Classification Report :

	precision	recall	f1-score	support
ac	0.87	0.66	0.75	673
apol	0.94	0.71	0.81	126
as	0.79	0.14	0.24	107
cfm_q	0.98	0.81	0.89	235
ch_q	0.57	0.13	0.21	63
cmd	1.00	0.90	0.95	352
cong	1.00	1.00	1.00	1069
cp	0.33	0.10	0.15	81
cp_q	1.00	0.06	0.11	18
dir	0.76	0.31	0.44	80
dny	0.88	0.77	0.82	636
gb	0.78	0.39	0.52	18
gt	0.86	0.53	0.66	83
inf	0.67	0.95	0.78	3535
inq_q	0.94	0.96	0.95	3812
instr	0.96	0.62	0.76	376
inv	0.99	0.93	0.96	182
op	0.84	0.50	0.62	721
otr_q	0.83	0.40	0.54	126
proh	0.93	0.85	0.89	267
req	0.94	0.86	0.90	1158
req_q	0.98	0.84	0.91	462
rev	0.68	0.22	0.33	118
s_i	1.00	0.99	1.00	3173
sug	0.99	0.87	0.92	599
thx	0.93	0.83	0.88	100
u	0.98	0.91	0.94	291
w	0.87	0.73	0.79	55
wn	1.00	0.07	0.13	14
accuracy			0.88	18530
macro avg	0.87	0.62	0.68	18530
weighted avg	0.89	0.88	0.87	18530

Figure 4.4 Classification Report for Polynomial Kernal

The classification report of Poly Kernel shown in Figure 4.4, which provides an evaluation of a model's performance across various classes, detailing precision, recall, F1-score, and support for each class. Overall, the model shows strong performance with an accuracy of 88%. The precision and recall values for most classes are high, indicating effective classification. Classes such as "cfm_q," "cmd," and "ch_q" achieved perfect precision and recall. However, certain classes like "as" and "dny" have lower recall scores, suggesting areas for improvement. The weighted average precision, recall, and F1-score are all high at 0.89, 0.88, and 0.87, respectively, highlighting the model's robust performance across the dataset.

4.4.4 Experimental Result of Naïve Bayes Classifier

The experiment involving the Naïve Bayes Classifier aimed to evaluate its effectiveness in classifying dialogue acts within a Myanmar language dataset. Two variations of the Naïve Bayes algorithm were utilized: MultinomialNB and BernoulliNB. The results underscored the classifiers' ability to capture relevant linguistic patterns, though the prevalence of common words indicated a potential area for improvement. Figures 4.5 and 4.6 present the classification reports for MultinomialNB and BernoulliNB, respectively.

```
Multinominal Naive Bayes
Accuracy on training Set :
0.8430227067283692
Accuracy on Testing Set :
0.7497571505666487
ကျွန်တော် ငွေလွှဲ ပါ
Predicted Target:  s_i
Actual Target:  s_i
နှမ်တစ် မြို့ ကို မ သွား ဘူး
Predicted Target:  dny
Actual Target:  dny
ဒါဘာ ဆိုင် လဲ ခိုင်
Predicted Target:  inq_q
Actual Target:  inq_q
ခင်ဗျား ခရီး တစ် ခု ထွက် မ လို့ လား
Predicted Target:  inq_q
Actual Target:  s_i
ကျွန်တော် ဒေါက်တာ မြင့်ရှိန် ပါ
Predicted Target:  s_i
Actual Target:  s_i
အင်း အဲ့ ဒါ လည်း မ ဆိုး ဘူး နော် အဲ့ ကို ပဲ သွား ကြ စို့
Predicted Target:  op
Actual Target:  u
```

```
Classification Report :
              precision    recall  f1-score   support

          ac       0.76      0.28      0.41       673
        apol       0.90      0.61      0.73       126
          as       0.90      0.24      0.38       107
      cfm_q       0.49      0.10      0.17       235
        ch_q       0.46      0.10      0.16        63
         cmd       0.72      0.77      0.74       352
        cong       0.99      1.00      1.00      1069
          cp       0.33      0.09      0.14        81
        cp_q       0.00      0.00      0.00        18
         dir       0.62      0.35      0.45        80
         dny       0.61      0.62      0.61       636
          gb       0.50      0.28      0.36        18
          gt       0.79      0.55      0.65        83
         inf       0.60      0.86      0.71      3535
       inq_q       0.75      0.92      0.82      3812
       instr       0.91      0.17      0.29       376
         inv       0.95      0.91      0.93       182
          op       0.61      0.16      0.26       721
       otr_q       0.61      0.17      0.27       126
        proh       0.58      0.12      0.20       267
         req       0.67      0.66      0.66      1158
       req_q       0.41      0.30      0.35       462
         rev       0.64      0.27      0.38       118
         s_i       1.00      1.00      1.00      3173
         sug       0.73      0.86      0.79       599
         thx       0.88      0.77      0.82       100
           u       0.99      0.32      0.48       291
           w       0.91      0.75      0.82        55
          wn       1.00      0.14      0.25        14

    accuracy                           0.75     18530
   macro avg       0.70      0.46      0.51     18530
weighted avg       0.75      0.75      0.72     18530
```

Figure 4.5 Classification Report for Multinomial Naïve Bayes

```
Bernoulli Naive Bayes
Accuracy on training Set :
0.8573105411567884
Accuracy on Testing Set :
0.8213707501349163
ကျွန်တော် ငွေလွှဲ ပါ
Predicted Target:  s_i
Actual Target:  s_i
နှမ်တစ် မြို့ ကို မ သွား ဘူး
Predicted Target:  dny
Actual Target:  dny
ဒါဘာ ဆိုင် လဲ ခိုင်
Predicted Target:  inq_q
Actual Target:  inq_q
ခင်ဗျား ခရီး တစ် ခု ထွက် မ လို့ လား
Predicted Target:  inq_q
Actual Target:  s_i
ကျွန်တော် ဒေါက်တာ မြင့်ရှိန် ပါ
Predicted Target:  s_i
Actual Target:  s_i
အင်း အဲ့ ဒါ လည်း မ ဆိုး ဘူး နော် အဲ့ ကို ပဲ သွား ကြ စို့
Predicted Target:  op
Actual Target:  u
```

```
Classification Report :
              precision    recall  f1-score   support

          ac       0.76      0.62      0.68       673
        apol       0.69      0.79      0.74       126
          as       0.58      0.60      0.59       107
      cfm_q       0.99      0.77      0.86       235
        ch_q       0.30      0.48      0.37        63
         cmd       0.65      0.91      0.76       352
        cong       1.00      1.00      1.00      1069
          cp       0.22      0.38      0.28        81
        cp_q       0.07      0.17      0.10        18
         dir       0.45      0.64      0.53        80
         dny       0.77      0.80      0.79       636
          gb       0.30      0.61      0.40        18
          gt       0.48      0.66      0.56        83
         inf       0.71      0.75      0.73      3535
       inq_q       0.92      0.86      0.89      3812
       instr       0.85      0.70      0.77       376
         inv       1.00      0.91      0.95       182
          op       0.61      0.51      0.56       721
       otr_q       0.38      0.52      0.44       126
        proh       0.94      0.83      0.88       267
         req       0.89      0.72      0.79      1158
       req_q       0.55      0.84      0.66       462
         rev       0.39      0.62      0.48       118
         s_i       0.99      1.00      0.99      3173
         sug       0.98      0.86      0.92       599
         thx       0.70      0.87      0.78       100
           u       0.95      0.87      0.91       291
           w       0.83      0.89      0.86        55
          wn       0.56      0.64      0.60        14

    accuracy                           0.82     18530
   macro avg       0.67      0.72      0.68     18530
weighted avg       0.84      0.82      0.83     18530
```

Figure 4.6 Classification Report for Bernoulli Naïve Bayes

4.4.5 Performance Comparison between SVM and NB

Table 4.3 illustrates the comparison of different classifiers for dialogue act recognition, highlighting their distinct performance characteristics across precision, recall, and F1-score.

Table 4.3 Dialogue Act Classification Score for Naïve Bayes and Different Kernel of SVM Classifier

	Precision	**Recall**	**F-score**
SVM Linear kernel	0.768	0.781	0.758
SVM RBF kernel	**0.796**	**0.815**	**0.787**
SVM Poly kernel	0.771	0.753	0.727
MultinomialNB	0.728	0.729	0.696
BernoulliNB	0.73	0.739	0.733

The SVM with an RBF kernel stands out, achieving the highest precision (0.796), recall (0.815), and F1-score (0.787), demonstrating its superior ability to classify and generalize across the dataset. The SVM with a linear kernel also performs well with a precision of 0.768, recall of 0.781, and an F1-score of 0.758, though it is slightly less effective than the RBF kernel. The SVM with a polynomial kernel shows a balanced but lower performance with a precision of 0.771, recall of 0.753, and an F1-score of 0.727. Among the Naive Bayes classifiers, BernoulliNB outperforms MultinomialNB, with a precision of 0.73, recall of 0.739, and an F1-score of 0.733, compared to MultinomialNB's precision of 0.728, recall of 0.729, and F1-score of 0.696. Overall, the SVM with the RBF kernel is identified as the most effective classifier, offering the best balance of precision, recall, and F1-score.

4.5 Summary

This chapter presents a comprehensive study on the application of heuristic and machine learning methods to recognize dialogue acts in the Myanmar language. Dialogue act recognition is a crucial component in natural language processing and understanding, as it involves identifying the functional role of each utterance in a dialogue. The study focuses on evaluating the performance of different machine

learning models, specifically Support Vector Machines (SVM) and Naive Bayes classifiers, in this context.

In the experiments conducted, various configurations of SVM were tested, including linear, radial basis function (RBF), and polynomial kernels. These different kernel functions influence how the SVM models interpret the data, potentially impacting their ability to correctly classify dialogue acts. Additionally, two types of Naive Bayes classifiers, MultinomialNB and BernoulliNB, were employed to determine their effectiveness in this task.

The performance of these models was assessed using standard evaluation metrics such as precision, recall, and F1-score. These metrics provide a nuanced understanding of each model's strengths and weaknesses by measuring the accuracy, completeness, and overall balance between precision and recall. The findings from the experiments indicated that the SVM with the RBF kernel outperformed the other models, demonstrating superior ability in recognizing dialogue acts in Myanmar dialogue.

These results offer significant insights into the suitability and efficiency of different machine learning techniques for dialogue act recognition in Myanmar. The study's findings underscore the potential of SVM, particularly with the RBF kernel, to enhance the accuracy of dialogue act recognition systems in the Myanmar language. This contributes valuable knowledge to the field of computational linguistics and natural language processing, especially in the context of less commonly studied languages.

CHAPTER 5

BIDIRECTIONAL LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK BASED MDAR

Dialogue Act Recognition (DAR) is a fundamental aspect of developing intelligent dialogue systems that can understand and process human language in a meaningful way. For languages with abundant resources like English, DAR has made significant advancements due to the availability of extensive annotated datasets and sophisticated NLP models. However, for low-resource languages such as Myanmar, DAR poses significant challenges. Myanmar language exhibits complex morphological and syntactical characteristics, making it difficult to develop effective language models with limited data. To address these challenges, employing Bidirectional Long Short-Term Memory Recurrent Neural Networks (Bi-LSTM RNNs) offers a promising solution.

Bi-LSTM RNNs are particularly suited for languages like Myanmar due to their ability to capture context from both past and future inputs within a sequence. This bidirectional context is crucial for accurately interpreting dialogue acts, as the meaning of an utterance often depends on both preceding and succeeding utterances. In Myanmar, where sentence structures and dependencies can be intricate, the ability of Bi-LSTM RNNs to understand and incorporate this dual context significantly enhances the performance of DAR systems. The importance of Bi-LSTM RNN-based DAR for Myanmar extends beyond technical performance. Developing robust dialogue systems for low-resource languages is essential for ensuring linguistic diversity in AI applications. This chapter provides a brief introduction to the Bi-LSTM RNN architecture utilized in this thesis. Experiments were conducted to compare Bi-LSTM RNN architectures with different word embedding models for MDAR.

5.1 Recurrent Neural Networks (RNNs)

RNNs are a class of artificial neural networks designed for sequential data, where the output from the previous step is fed as input to the current step. This chain-like nature is particularly effective for tasks involving time series or language data. However, standard RNNs suffer from the vanishing gradient problem, making them

inefficient for long sequences. Long Short-Term Memory (LSTM) networks address this issue by introducing memory cells that can retain information for long periods.

5.1.1 Theoretical Foundation of LSTM

An LSTM network, introduced by Hochreiter and Schmidhuber in 1997 [73], is designed to overcome the vanishing gradient problem in traditional RNNs. The LSTM unit consists of a cell state C_t, which carries information across time steps, and three types of gates: the input gate i_t, the forget gate f_t, and the output gate o_t. These gates control the flow of information in and out of the cell. The mathematical formulation of an LSTM unit is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5.1}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5.2}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5.3}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5.4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5.5}$$

$$h_t = o_t * \tanh(C_t) \tag{5.6}$$

Here, σ denotes the sigmoid function, W are weight matrices, b are biases, i_t is the input gate, f_t is the forget gate, o_t is the output gate, C_t is the cell state, and h_t is the hidden state at time step t. Figure 5.1 gives the fundamental structure of LSTM unit.

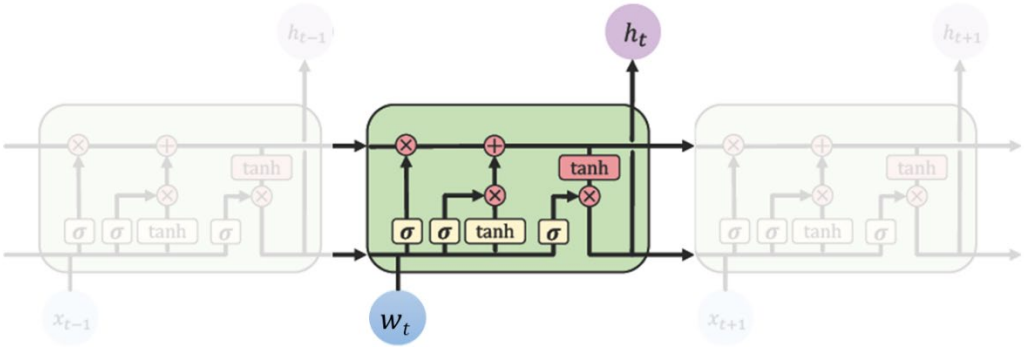


Figure 5.1 Schematic of LSTM Unit (Figure Source: [86])

5.1.2 Bidirectional LSTM (Bi-LSTM)

While LSTMs can capture long-term dependencies, they process data in a unidirectional manner, either from past to future or vice versa. This unidirectional nature might limit the context captured, especially in language processing where the meaning of a word often depends on both preceding and succeeding words. Bidirectional LSTM (Bi-LSTM) networks [62] overcome this limitation by processing the sequence in both forward and backward directions, thus capturing richer contextual information.

A Bi-LSTM consists of two LSTMs: one processes the sequence from start to end (forward LSTM), and the other processes it from end to start (backward LSTM). The outputs of both LSTMs are concatenated at each time step. Mathematically, this can be expressed as:

$$\overrightarrow{h_t} = \text{LSTM}_{\text{forward}}(x_t, \overrightarrow{h_{t-1}}) \tag{5.7}$$

$$\overleftarrow{h_t} = \text{LSTM}_{\text{backward}}(x_t, \overleftarrow{h_{t+1}}) \tag{5.8}$$

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{5.9}$$

where $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are the hidden states of the forward and backward LSTMs, respectively, and h_t is the concatenated hidden state.

5.2 Word Embedding Model

Word embedding is a form of natural language processing (NLP) that uses vectors to map text from a corpus. To put it another way, it is a learned representation that lets two words share the same representation. The distributed representation of a text (words and documents) is a significant breakthrough for improved performance in NLP-related problems. By retaining context word similarity and using low-dimensional vectors, word embedding provides a more efficient and expressive representation. Many fields, including semantic analysis, dialogue act classification, and machine translation, now use word embedding.

Word2Vec [80] is the well-known models of word representation methods. It converts each word into a fixed-length vector, which can better express the analogy and similarity relationship between different words. It includes continuous bag-of-words

(CBOW) and continuous skip-gram models for learning word embeddings (as words are embedded in a vector space) directly from unannotated text distributions. Conditional probabilities, which can be thought of as predicting some words based on some of their surrounding words in corpora, are used in their training for semantically meaningful representations. The CBOW model predicts a target word using the embeddings of its context words, whereas the skip-gram model uses the embedding of a target word to predict surrounding words. Both skip-gram and continuous bag of words models are self-supervised because supervision is derived from data without labels.

5.2.1 Continuous Bag of Words (CBOW) Model

The continuous bag of words (CBOW) model differs from the skip-gram model in that it assumes that a center word in the text sequence is generated based on the context words surrounding it. When the sentence “Let’s travel to Kalaw” translate to Myanmar, which consider the words “ကလော”, “ကို”, “ခရီးသွား”, “ကြ”, and “စို့”, with “ခရီးသွား” as the center word, and the context window size of 2, the conditional probability of generating the center word “ခရီးသွား” based on the context words “ကလော”, “ကို”, “ကြ”, and “ စို့” is considered in the continuous bag of words model.

$$P\ (\text{“ခရီးသွား”} \mid \text{“ကလော”}, \text{“ကို”}, \text{“ကြ”}, \text{“စို့”}) \tag{5.10}$$

The conditional probability is calculated by averaging the context word vectors because the continuous bag of words model contains multiple context words. Denote by $u_i \in \mathbb{R}^d$ and $v_i \in \mathbb{R}^d$ are the two vectors of any word in the dictionary with index i when applied as a center word and a context word, respectively. The conditional probability of producing any center word ω_c (with the dictionary index c) given its surrounding context words $\omega_{o_1}, \ldots, \omega_{o_{2m}}$ (with the dictionary index $o_1, \ldots, o_{2m}$) can be modeled as follow,

$$P(\omega_c \mid \omega_{o_1}, \ldots, \omega_{o_{2m}}) = \frac{exp\left(\frac{1}{2m}\, u_c^T\, (v_{o_1} + \cdots, + v_{o_{2m}})\right)}{\sum_{i \epsilon V} exp\left(\frac{1}{2m}\, u_i^T\, (v_{o_1} + \cdots, + v_{o_{2m}})\right)} \tag{5.11}$$

For brevity, let $\mathcal{W}_o = \{\omega_{o_1}, \ldots, \omega_{o_{2m}}\}$ and $\bar{v}_o = (v_{o_1} + \cdots, +v_{o_{2m}})/(2m)$. Then the above equation can be simplified as

$$P(\omega_c \mid \mathcal{W}_o) = \frac{exp(u_c^T \bar{v}_o)}{\sum_{i \epsilon \mathcal{V}} exp(u_i^T \bar{v}_o)} \tag{5.12}$$

Given a text sequence of length T, where $w^{(t)}$ denotes the word on each time step t. The probability of generating all center words given their context words is the likelihood function of the continuous bag of words model for context window size m:

$$\prod_{t=1}^{T} P\left(\omega^{(t)} \mid \omega^{(t-m)}, \ldots, \omega^{(t-1)}, \omega^{(t+1)}, \ldots, \omega^{(t+m)}\right) \tag{5.13}$$

Unlike the skip-gram model, the continuous bag of words model typically uses context word vectors as word representations in dialogue act recognition.

5.2.2 Skip-Gram Model

The skip-gram model assumes that a word in a text sequence can be used to generate the words around it. In the same text sequence “Let’s travel to Kalaw” includes the words “ကလော”, “ကို”, “ခရီးသွား”, “ကြ”, and “စို့” in Myanmar. Set the size of the context window to 2 and enter “travel” as the center word. The skip-gram model considers the conditional probability of generating the context words “ကလော”, “ကို”, “ကြ”, and “စို့”, which are no more than two words away from the center word.

$$P\left(\text{“ကလော”}, \text{“ကို”}, \text{“ကြ”}, \text{“စို့”} \mid \text{“ခရီးသွား”}\right) \tag{5.14}$$

Assume that the context words are generated independently of the center word (i.e., conditional independence). The preceding conditional probability can be rewritten as follows.

$$P(\text{“ကလော”}|\text{”ခရီးသွား”}).P(\text{“ကို”}|\text{”ခရီးသွား”}).P(\text{“ကြ”}|\text{”ခရီးသွား”}).P\left(\text{“စို့”}|\text{”ခရီးသွား”}\right) \tag{5.15}$$

Each word in the skip-gram model has two d-dimensional-vector representations for calculating conditional probabilities. $v_i \in \mathbb{R}^d$ and $u_i \in \mathbb{R}^d$ represent the two vectors of any word with index *i* in the dictionary when it is used as a center word and a context word, respectively (meanings are switched in the continuous bag of words

model). The conditional probability of generating any context word ω_o (with index o in the dictionary) given the center word ω_c (with index c in the dictionary) can be modelled using a softmax operation on vector dot products:

$$P(\omega_o \mid \omega_c) = \frac{exp(u_o^T v_c)}{\sum_{i \epsilon \mathcal{V}} exp(u_i^T v_c)} \tag{5.16}$$

where $V = \{0,1,\ldots,|V|-1\}$ represents the vocabulary index. Given a text sequence of length T, where w represents the word at time step t. Assume that context words are generated independently of any center word. The skip-gram model is likelihood function for context window size m is the probability of generating all context words given any center word:

$$\prod_{t=1}^{T} \prod_{-m \le j \le m, j \ne 0} P(\omega^{(t+j)} \mid \omega^{(t)}) \tag{5.17}$$

where any time step that is less than 1 or greater than T can be omitted.

The center word vectors of skip-gram model are commonly used as word representations in dialogue act recognition. Figure 5.2 shows the comparison of the CBOW and Skip-gram models architecture for an example sentence.

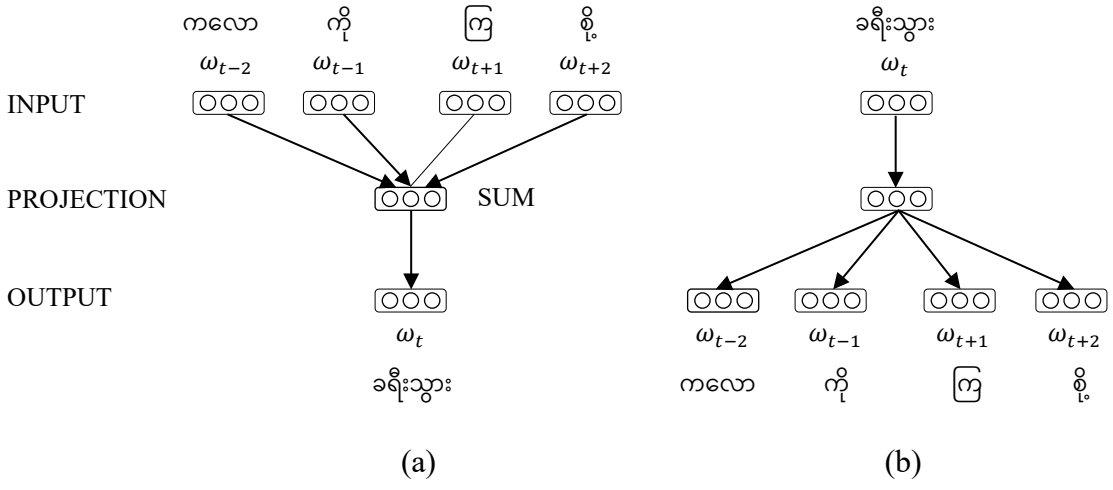
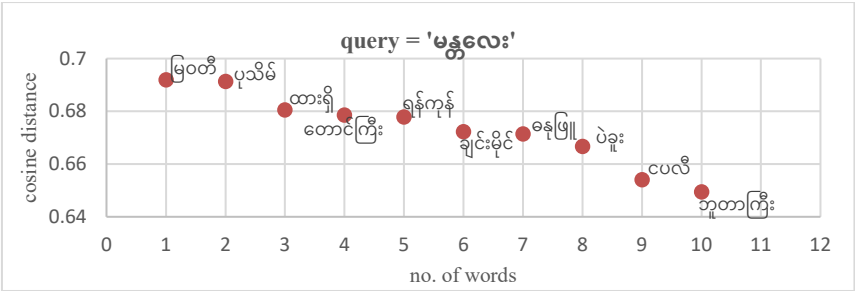
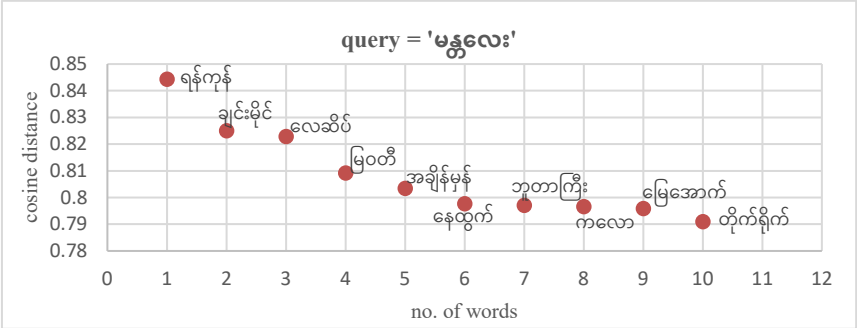


Figure 5.2 Model Architectures: (a) CBOW and (b) Skip-gram

Figure 5.3 shows examples of the list of nearest neighbors of words that can be generated using 300-dimensional of Skip-gram and CBOW word vector model.



(a)



(b)

Figure 5.3 Comparison of the Nearest Neighbors Words on (a) Skip-gram and (b) CBOW Model

5.3 Bi-LSTM RNN based MDAR

Figure 5.4 illustrates the architecture of a Bi-LSTM (Bidirectional Long Short-Term Memory) network for sequence labeling tasks, such as dialogue act recognition. The model processes word tokens w_{k1}, $w_{k2,}$ ..., $w_{ki,}$ ..., w_{kt} corresponding to each time step u_t. Each word token is first mapped to a dense vector in the embedding layer, capturing semantic information. These embeddings are then fed into the Bi-LSTM layer, which consists of forward and backward LSTM cells that capture contextual information from both directions in the sequence. The outputs from the Bi-LSTM layer at each time step $\overleftarrow{h_t}$ and $\overrightarrow{h_t}$ are concatenated to form the final hidden state h_t. This hidden state is then passed to the output layer, which generates the final prediction $\hat{y}_t$ for each time step. The architecture leverages the bidirectional nature of the LSTM to utilize past and future context, enhancing the model's performance on tasks involving sequential data.

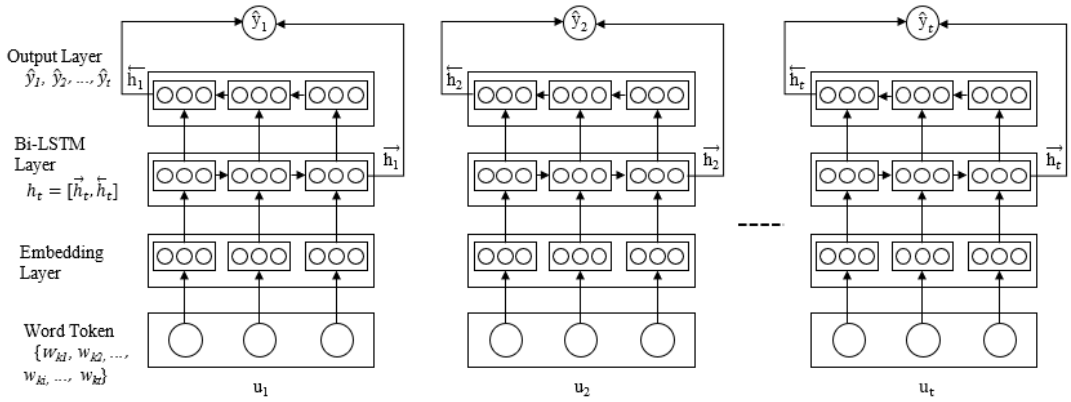


Figure 5.4 An illustration of the proposed Bi-LSTM based RNN model

5.3.1 Experiment

To implement the neural network model training, the TensorFlow [88] and Keras [87] libraries are utilized for building and training the neural network due to their flexibility and robustness in handling various feature inputs and output structures, enabling precise customization to fit the specific requirements of the dialogue act recognition task. The experiments are conducted on an Nvidia Tesla K80 GPU, which provides the necessary computational power to handle the complex operations involved in training deep learning models.

Table 5.1 Experiments of Best Fine Tuned Hyperparameters for BI-LSTM RNN Model

Parameters	Values
Embedding size	128
Hidden layer size	512
Max-sequence length	40
Learning rate	0.001
Dense	29
Batch-size	512
Number of epochs	100

Different parameter settings, such as batch size, learning rate, and the number of epochs, are tested to optimize model performance. These variations significantly impact the training duration, as each configuration can affect the convergence rate and overall computational demands. By leveraging TensorFlow's robust capabilities and the high-performance Nvidia Tesla K80 GPU, the experiments aim to efficiently train and evaluate the neural network models, ensuring thorough exploration and fine-tuning of

the optimal parameters for accurate dialogue act recognition in the Myanmar language. Tabel 5.1 shows the hyperparameters tuning lists of the Bi-LSTM model architecture.

The MmTravel corpus, as detailed in Chapter 3, served as the foundational dataset for modeling dialogue act recognition in the Myanmar language. This corpus was meticulously constructed to capture a wide range of conversational contexts relevant to travel, providing a robust dataset for training and evaluation.

To assess the effectiveness of the models, three distinct approaches were evaluated: a baseline Bi-LSTM RNN model, a Bi-LSTM RNN model enhanced with Skip-gram Word2Vec embeddings, and a Bi-LSTM RNN model utilizing CBOW Word2Vec embeddings. The baseline Bi-LSTM RNN model leverages the recurrent neural network architecture to capture sequential dependencies in the dialogue data, aiming to recognize the underlying dialogue acts based on the given input sequences. The second model incorporates Skip-gram Word2Vec embeddings, which are trained to predict surrounding words within a window of context words, thus capturing more nuanced semantic relationships and potentially enhancing the model's understanding of word meanings and their contextual usage. The third model employs CBOW (Continuous Bag of Words) Word2Vec embeddings, which predict a word based on its context, offering another perspective on capturing semantic information from the text.

Each of these models was rigorously tested using the MmTravel corpus to determine their performance in recognizing dialogue acts. By comparing the results across these models, the study aims to identify the most effective approach for dialogue act recognition in the Myanmar language.

5.3.2 Evaluation on the Bi-LSTM RNN model

The Bi-LSTM RNN model's evaluation metrics, with a loss of 0.6694 and an accuracy of 0.8990, demonstrate robust performance. The low loss indicates effective error minimization, while the high accuracy shows the model's strong predictive capability. These results suggest the model has learned the data patterns well and can generalize reliably to new data, making it suitable for practical applications.

Figure 5.5 shows the comparison of the model accuracy and loss over the first 100 epochs of the Bi-LSTM RNN model reveals significant trends. The accuracy graph shows a rapid increase in the initial epochs, reaching around 0.9 by epoch 10, and

gradually approaching 1.0 as training progresses, indicating that the model quickly learns to make correct predictions and continues to improve at a slower rate thereafter. On the other hand, the loss graph demonstrates a steep decline in the initial epochs, dropping from a high of around 1.6 to below 0.4 by epoch 10, and then tapering off and stabilizing with minor fluctuations for the remaining epochs. This indicates that the model rapidly reduces its prediction errors initially and then fine-tunes its performance to minimize loss further. Overall, the model shows efficient learning and convergence, achieving high accuracy and low loss within the 100 epochs.

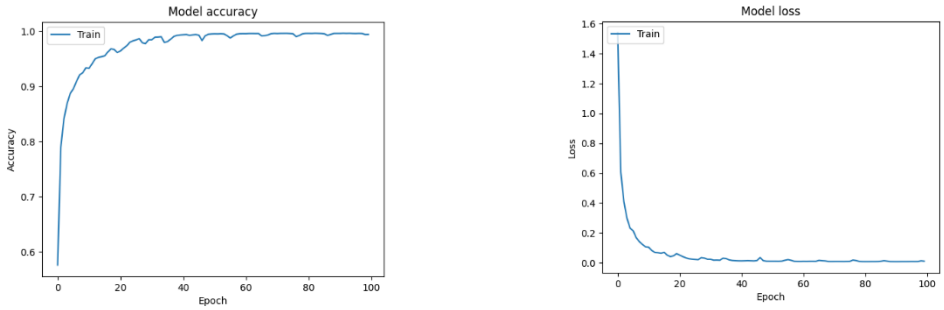


Figure 5.5 Accuracy and Loss for the 100 Epochs on a Bi-LSTM RNN Model

5.3.3 Evaluation on the Bi-LSTM RNN model with word2vec

The comparison between the Skip-gram and CBOW models reveals a clear performance advantage for the Skip-gram approach over 300 epochs. Skip-gram achieves an accuracy of 84.48%, a precision of 87.25%, a recall of 84.48%, and an F1 score of 85.40%. These metrics indicate that the Skip-gram model is highly effective in correctly predicting and capturing relevant features. In contrast, the CBOW model shows significantly lower performance, with an accuracy of 58.21%, a precision of 62.95%, a recall of 54.21%, and an F1 score of 52.65%. The substantial gap between the two models across all metrics suggests that Skip-gram is more proficient in learning meaningful representations from the data, making it a superior choice for tasks requiring high accuracy and reliable feature extraction.

Figure 5.6 describes the comparison of the Skip-gram and CBOW models based on their accuracy and loss over 100 epochs highlights the superior performance of the Skip-gram model. In the context of Bi-LSTM RNN for dialogue act recognition, it is crucial to assess these metrics to choose the best model. The Skip-gram model demonstrates a rapid increase in accuracy, plateauing around 0.95 by the 100th epoch,

indicating efficient learning of word relationships early on. The slight fluctuations towards the end suggest potential overfitting, but the high final accuracy indicates that it captures detailed semantic information, crucial for understanding context in dialogue act recognition. The loss curve supports this, with a sharp early drop and a stable, low value, indicating effective error minimization and strong generalization.

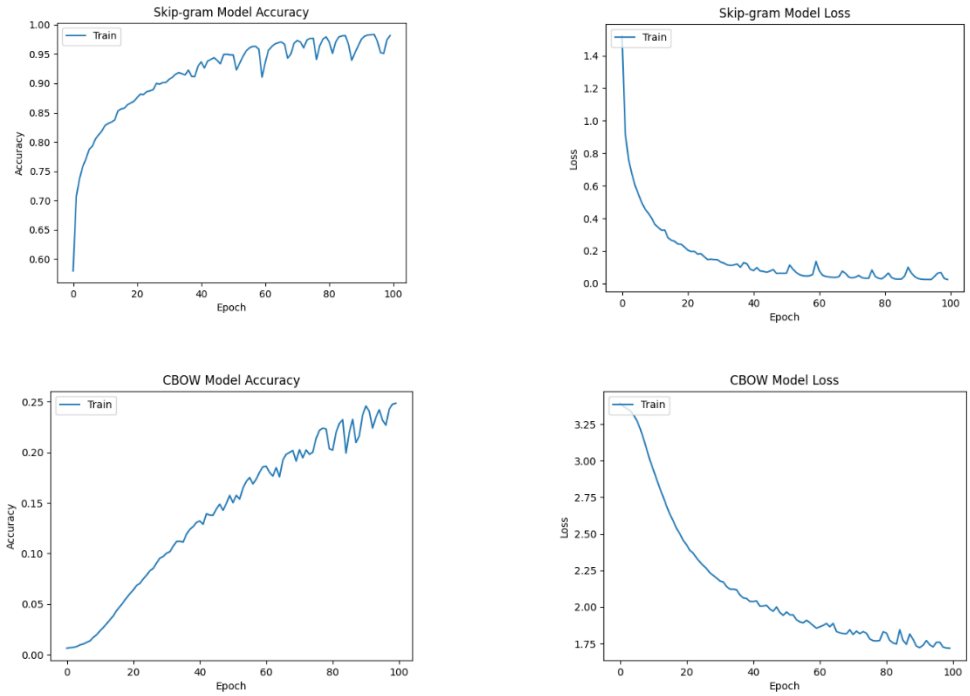


Figure 5.6 Accuracy and Loss for the 100 Epochs on a Bi-LSTM RNN with Word2Vec Embedding

Conversely, the CBOW model shows a slower, steady increase in accuracy, reaching around 0.22 after 100 epochs. This suggests it takes longer to learn word relationships, and the lower final accuracy implies it captures less semantic detail, limiting its effectiveness for dialogue act recognition. The loss curve, starting higher and decreasing gradually, indicates that CBOW requires more epochs to converge and struggles with generalization compared to Skip-gram.

When comparing the two models, the Skip-gram model clearly outperforms the CBOW model in both accuracy and loss metrics. The Skip-gram model's ability to rapidly learn and achieve high accuracy while maintaining a low and stable loss makes it a more suitable candidate for word embedding tasks in a Bi-LSTM RNN designed for dialogue act recognition. The high-quality embeddings produced by the Skip-gram model can enhance the Bi-LSTM RNN's performance, allowing it to better understand

and predict dialogue acts by leveraging the detailed semantic relationships embedded in the word vectors. On the other hand, the CBOW model, with its slower learning and lower final accuracy, might not provide the same level of detail and precision, potentially limiting the effectiveness of the Bi-LSTM RNN in capturing the complexities of dialogue acts. Therefore, for dialogue act recognition tasks that demand high accuracy and precise semantic understanding, the Skip-gram model is the preferred choice for word embedding in conjunction with a Bi-LSTM RNN.

5.3.4 Evaluating the Performance of Neural Networks Versus Baseline SVM Model

In this section, the performance of neural networks is assessed by comparing to a baseline Support Vector Machine (SVM) model. Many fields of artificial intelligence have been revolutionized by neural networks, particularly deep learning architectures, through the provision of powerful tools for pattern recognition and complex data analysis. By contrast, a staple in machine learning for decades has been SVMs, known for their robustness and effectiveness in various classification tasks. The strengths and weaknesses of each approach are aimed to be determined within the context of our specific application. Metrics such as accuracy, precision, recall, and computational efficiency will be compared to provide a comprehensive analysis. Not only will the advancements brought by neural networks be highlighted by this comparison, but scenarios where traditional models like SVM might still be preferable will also be underscored. Insights that guide the choice of model for future projects and applications are aimed to be drawn through this rigorous evaluation.

Table 5.2 Average Precision, Recall, F1-Score, and Overall Accuracy of the Bi-LSTM RNN Model Compared with Other Models on the MmTravel Dataset

Model	%			
	Precision	***Recall***	***F1***	***Accuracy***
Bi-LSTM RNN	**85.25**	72.70	**78.65**	**80.70**
LSTM RNN	76.61	57.58	65.74	72.03
RNN	59.02	62.07	60.50	68.00
SVM	66.90	**75.35**	71.59	77.80

The performance comparison of the various models in Table 5.2 shows significant differences in their ability to classify data accurately. The Bi-LSTM RNN outperforms the others with the highest precision (85.25%), F1 score (78.65%), and

accuracy (80.70%), indicating a strong balance between precision and recall (72.70%). The LSTM RNN, while maintaining a reasonable precision of 76.61%, falls short in recall (57.58%) and F1 score (65.74%), resulting in a lower accuracy of 72.03%. The basic RNN model exhibits the lowest precision (59.02%) but a relatively higher recall (62.07%), leading to an F1 score of 60.50% and an accuracy of 68.00%. The SVM model achieves a balanced performance with a precision of 66.90%, the highest recall (75.35%), an F1 score of 71.59%, and an accuracy of 77.80%, making it the second most effective model in this comparison. Overall, the Bi-LSTM RNN is the most robust model, with SVM being a competitive alternative, particularly in recall.

Table 5.3 Precision (P), Recall (R), and F1-Score (F1) Classification Scores for Bi-LSTM Model Across Training, Validation, and Testing Sets

act	Train Set			Validation Set			Test Set		
	P	***R***	***F1***	***P***	***R***	***F1***	***P***	***R***	***F1***
inf	76.73	50.29	60.76	77.06	49.31	60.14	78.42	57.91	66.62
ac	11.67	4.14	6.11	14.53	5.03	7.47	55.24	90.62	68.64
apol	51.89	85.94	64.71	46.4	90.62	61.38	96	75	84.21
cp	10.71	31.25	15.96	9.09	25	13.33	6.52	6.25	6.38
cong	99.44	100	99.72	99.44	99.81	99.62	99.44	100	99.72
dny	69.31	88.27	77.65	70.83	83.06	76.46	73.17	78.18	75.59
op	51.97	66.85	58.48	50.63	67.7	57.93	34.38	80.06	48.1
rev	21.58	69.49	32.93	18.58	71.19	29.47	20.62	33.9	25.64
thx	64.29	93.75	76.27	64.38	97.92	77.69	70.59	100	82.76
w	63.64	84	72.41	58.33	84	68.85	35.38	92	51.11
cmd	89.8	94.62	92.15	95.65	94.62	95.14	96.17	94.62	95.39
dir	32.2	51.35	39.58	29.49	62.16	40	58.33	37.84	45.9
instr	45.76	62.43	52.81	46.35	62.43	53.2	44.57	66.47	53.36
inv	89.22	98.91	93.81	94.79	98.91	96.81	98.91	98.91	98.91
proh	84.62	95.28	89.63	82.31	95.28	88.32	67.21	96.85	79.35
req	86.5	95.58	90.82	87.93	95.58	91.6	95.81	84.49	89.79
sug	85.62	82.78	84.18	84.8	83.11	83.95	85.96	83.11	84.51
u	78.57	97.95	87.2	79.01	97.95	87.46	91.08	97.95	94.39
wn	28.2	27.61	29.14	30.28	32.8	34.2	25.88	68.75	37.61
ch_q	31.03	56.25	40	31.82	65.62	42.86	76.47	40.62	53.06
cp_q	22.6	24.19	26.94	24.71	25.68	27.35	88.28	92.62	90.4
cfm_q	77.93	92.62	84.64	86.26	92.62	89.33	94.17	92.62	93.39
inq_q	93.42	82.69	87.73	94.17	83.65	88.6	88.07	92.63	90.29
otr_q	63.38	68.18	65.69	60	72.73	65.75	30.5	65.15	41.55
req_q	46.68	87.19	60.81	48.15	85.95	61.72	66.78	78.93	72.35
gb	9.09	7.69	8.33	16.67	7.69	10.53	30.77	27.27	28.92
gt	36.59	34.09	35.29	32.14	40.91	36	14.29	13.64	13.95
s_i	97.34	98.07	97.7	97.7	98.07	97.89	98.87	97.94	98.4
as	24.79	50.88	33.33	24.19	52.63	33.15	34.85	40.35	37.4

Table 5.3 shows the comparison of precision (P), recall (R), and F1 scores across the training, validation, and test sets reveals varying performance for different categories. For instance, "inf" maintains relatively consistent P, R, and F1 scores across all sets, with an F1 score improvement in the test set 66.62% compared to the train 60.76% and validation 60.14% sets, indicating robust generalization. Conversely, "ac" shows significant improvement in the test set, with an F1 score jumping to 68.64% from much lower values in the train 6.11% and validation 7.47% sets, suggesting a substantial enhancement in model handling of this action type during evaluation.

In contrast, some action types exhibit significant disparities between the datasets. The "cp" action type performs poorly across all sets, with F1 scores remaining low, particularly in the test set at 6.38%, indicating consistent difficulties in accurately predicting this category. On the other hand, "apol" shows a dramatic increase in the test set F1 score to 84.21%, significantly higher than its performance in the train 64.71% and validation 61.38% sets, reflecting excellent model generalization for this type. Actions like "cong" maintain near-perfect scores across all sets, with F1 scores hovering around 99.72%, demonstrating high accuracy and reliability in predicting this action type. Other action types, such as "req_q" and "s_i," also exhibit strong performance with high F1 scores across all datasets, highlighting the model's effectiveness in these categories. The comparison of these metrics helps identify areas of strength and opportunities for improvement in the models.

5.4 Summary

According to the experimental results, it has been revealed that bidirectional LSTM-based deep neural models are powerful for Myanmar Dialogue Act Recognition. The Bi-LSTM RNN model demonstrated superior performance in terms of precision, recall, F1-score, and overall accuracy when compared to other models on the MmTravel dataset. This indicates the effectiveness of Bi-LSTM RNNs in handling the complex morphological and syntactical characteristics of the Myanmar language, making them a robust choice for dialogue act recognition tasks.

CHAPTER 6
CONCLUSION AND FUTURE WORK

This chapter provides a comprehensive summary of the research work, highlighting both the advantages and limitations of the proposed dissertation. The primary contribution of this research is the pioneering evaluation of machine learning and deep neural network architectures for Myanmar Dialogue Act Recognition. This study marks the first instance of such an evaluation in the context of the Myanmar language, setting a foundation for future advancements in this field. A significant part of the research involved the manual development and proposal of a dialogue-tagged corpus for the Myanmar language. This corpus serves as a critical resource for training and evaluating dialogue act recognition models, addressing the lack of existing annotated datasets in this language.

The research showcases the implementation and comparison of various machine learning techniques, including Support Vector Machines (SVM) with different kernel functions, Naïve Bayes (NB) classifier and deep neural network architectures, to determine their effectiveness in recognizing dialogue acts in the Myanmar language. The evaluation results provide valuable insights into the strengths and weaknesses of each approach, guiding future research directions and potential improvements.

Overall, this chapter encapsulates the innovative efforts and findings of the research, emphasizing its role in advancing natural language processing for the Myanmar language while also acknowledging the challenges and areas for further exploration.

6.1 Advantages of the System

The Myanmar Dialogue Act Recognition system, utilizing Bi-LSTM RNN with word2vec embedding models, enhances communication efficiency and accuracy across various sectors. It improves automated customer service by accurately identifying and categorizing user intentions, promotes personalized learning in educational settings, and aids in the precise interpretation of patient symptoms in healthcare, leading to more effective interventions.

The use of Bi-LSTM RNN with word2vec embeddings provides significant advantages over traditional machine learning methods such as SVM and NB. The Bi-LSTM RNN captures long-term dependencies and contextual information in dialogue, crucial for understanding the nuances of natural language. Leveraging word2vec embeddings allows the model to benefit from dense vector representations of words, enhancing its ability to discern semantic similarities and relationships. This approach contrasts with SVM and NB, which rely on predefined features and often struggle with the inherent complexity and variability of dialogue data. Compared to the SVM baseline, the Bi-LSTM RNN with word2vec embeddings demonstrates superior performance in handling sequential data and contextual nuances, leading to more accurate and robust dialogue act recognition. Overall, the Myanmar Dialogue Act Recognition system enhances the quality and effectiveness of communication across multiple sectors, fostering better user experiences and operational efficiencies.

6.2 Limitations of the System

The Myanmar Dialogue Act Recognition (MDAR) system, while a promising development in natural language processing (NLP) for the Burmese language, faces several limitations that impede its effectiveness and broader application. One primary limitation is the lack of extensive and high-quality annotated datasets in the Burmese language. Compared to languages like English, Burmese has significantly fewer resources available for training and testing NLP models. This scarcity hinders the development of robust dialogue act recognition systems that can accurately interpret and respond to diverse conversational contexts.

A significant challenge for the MDAR system is the complexity of the Burmese language, with its rich system of tones, honorifics, and contextual nuances, poses difficulties for current NLP technologies that are primarily designed for languages with simpler grammatical structures. These linguistic features can dramatically alter the meaning of a sentence, making it challenging for MDAR systems to accurately recognize and classify dialogue acts, which can lead to misinterpretations and errors in dialogue management.

The word "ဟား ဟား" (Hah Hah) is inherently an expressive speech act, primarily used to convey laughter and amusement. However, in this work, it is assigned to the category of opinion speech acts due to the desire to avoid creating new dialogue acts. This classification creates a significant limitation, as it forces an expressive, emotional reaction into a framework meant for conveying beliefs, judgments, or evaluations. The expressive nature of "ဟား ဟား" does not fit neatly into the opinion category, potentially leading to misunderstandings or misinterpretations of the speaker's intent. This approach underscores a limitation in the methodology, as it sacrifices the nuance and accuracy of communication for the sake of simplicity, highlighting the challenges of rigidly categorizing speech acts in dialogue systems.

6.3. Future Work

The future work of Myanmar Dialogue Act Recognition aims to leverage contextual information to enhance the accuracy and relevance of its applications. A primary focus will be on utilizing the dialog acts from previous utterances to provide a more comprehensive understanding of ongoing conversations. By integrating historical dialogue context, the system can have better predict and interpret the current dialogue acts, leading to more nuanced and accurate recognition. This approach ensures that the system not only considers isolated statements but also understands the flow and progression of the conversation, thereby improving its ability to respond appropriately.

Additionally, developing a multitask model that combines intent recognition and slot filling will significantly improve natural language understanding for the Myanmar language by identifying user intent and extracting relevant information simultaneously. This integration allows for more efficient parsing of complex utterances, leading to more seamless and sophisticated interactions. The multitask approach enhances the robustness and versatility of the Dialogue Act Recognition system, enabling it to handle diverse conversational scenarios from simple queries to complex dialogues. By addressing multiple aspects of natural language understanding, the model adapts to varying contexts and user needs, ensuring effective communication and user satisfaction. This marks significant progress in developing advanced, context-aware conversational agents for the Myanmar language.

Author's Publications

1. Sann Su Su Yee, ChenChen Ding, Khin Mar Soe, Masao Utiyama, and Eiichiro Sumita, "Modifying NOVA-annotated Myanmar Data to Universal Part-of-Speech Tagset". In Proceedings of the 17th International Conference on Computer Applications (ICCA 2019), Yangon, Myanmar, pp. 230-237, February 27-28, 2019.
2. Chenchen Ding, Sann Su Su Yee, Win Pa Pa, Khin Mar Soe, Masao Utiyama, and Eiichiro Sumita. 2020. A Burmese (Myanmar) Treebank: Guideline and Analysis. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 19, 3, Article 40, January 2020.
3. Sann Su Su Yee, Khin Mar Soe, Ye Kyaw Thu, "Myanmar Dialogue Act Recognition (MDAR)". 2020 IEEE Conference on Computer Applications (ICCA), Yangon, Myanmar, February 2020, pp. 1-6.
4. Sann Su Su Yee, Khin Mar Soe, "Myanmar Dialogue Act Recognition Using Bi-LSTM RNN", 23rd Conference of the Oriental COCOSDA, (Oriental COCODSA 2020), Yangon, Myanmar, November 2020.
5. Sann Su Su Yee, Khin Mar Soe, "Joint Intent Detection and Slot Filling with BERT in Burmese", 2022 International Conference on Communication and Computer ResearchAt: Sookmyung Women's University, Seoul, Korea.
6. Hsu Myat Mo, Sann Su Su Yee, Khin Mar Soe, "DIET Architecture for Users' Comments in Myanmar Language", 2023 International Conference on Communication and Computer ResearchAt: Hotel Interciti, Daejeon, Korea.
7. Sann Su Su Yee, Khin Mar Soe, "Exploring BERT-based Encoders for Sequence Classification and Multi-task Learning in Dialogue Acts and Joint Intent-Slot Filling", Indian Journal of Computer Science and Engineering (IJCSE), Engg Journals Publications, 2024.

BIBLIOGRAPHY

[1] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional lstm and other neural network architectures,” NEURAL NETWORKS, pp. 5–6, 2005.

[2] A. H Anderson et al., “The HCRC map task corpus”, in: Language and speech 34.4, pp. 351–366, 1991.

[3] A. Janin et al. “The ICSI Meeting Corpus”, in: Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP’03). 2003 IEEE International Conference on. Vol. 1. IEEE, pp. I–364–I–367, 2003.

[4] A. Margolis, K. Livescu, and M. Ostendorf, “Domain adaptation with unlabeled data for dialog act tagging”, in: Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing. Association for Computational Linguistics, pp. 45–52.

[5] A. McCallum, and K. Nigam, “A comparison of event models for Naive Bayes text classification”, in: *AAAI-98 Workshop on Learning for Text Categorization* (pp. 41–48). CA: AAAI Press, 1998.

[6] A. Schmitt, S. Ultes, W. Minker, “A Parameterized and Annotated Spoken Dialog Corpus of the CMU Let's Go Bus Information System”, in: N. C. C. Chair), K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Eight International Conference on Language Resources and Evaluation, LREC'12, European Language Resources Association (ELRA), Istanbul, Turkey, 2012.

[7] A. Simpson, “The grammaticalization of clausal nominalizers in Burmese”, In Rethinking Grammaticalization, ed. María José López-Couso and Elena Soane, 265–288. John Benjamins Publishing Company, 2008

[8] A. Stolcke et al., “Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech”, in: Computational Linguististics 26.3, pp. 339–373, 2000.

[9] B. Gambäck, F. Olsson, and O. Täckström , “Active learning for dialogue act classification”, in: INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association. International Speech Communication Association, pp. 1329–1332.

[10] B. Krause, L. Lu, I. Murray, and S. Renals, “Multiplicative lstm for sequence modelling,” 2016.

[11] B. Prachya, S. Thepchai, “Technical report for the network-based asean language translation public service project”, Online Materials of Networkbased ASEAN Languages Translation Public Service for Members, NECTEC, 2013.

[12] B. Samei et al., “Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, June 5-9, 2014. Proceedings”, in: Springer. Chap. Context-Based Speech Act Classification in Intelligent Tutoring Systems, pp. 236–241.

[13] C. Bothe, C.Weber, S. Magg, and S.Wermter, “A context-based approach for dialogue act recognition using simple recurrent neural networks”, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), (Miyazaki, Japan), European Languages Resources Association (ELRA), May 2018.

[14] C. Cortes, V. Vapnik, “Support vector networks. Machine Learning”, 20(3), 273–297, 1995.

[15] C. D Manning, H. Schütze, and P. Raghavan “Introduction to information retrieval”, Cambridge university press, 2008.

[16] C. D. Manning, M. Surdeanu, J.Bauer, J. R. Finkel, S. Bethard, and D. McClosky, “The stanford corenlp natural language processing Toolkit”., ACL (System Demonstrations), pp. 55-60, 2014.

[17] C. Moldovan, V. Rus, and A. C Graesser, “Automated Speech Act Classification For Online Chat”, in: 22nd Midwest Artificial Intelligence and Cognitive Science Conference (MAICS 2011). MAICS, pp. 23–29.

[18] D. Das et al. (2014). “Frame-semantic parsing”, in: Computational Linguistics 40.1, pp. 9–56.

[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Neurocomputing: Foundations of research,” in:. Learning Representations by Back-propagating Errors, pp. 696–699, Cambridge, MA, USA: MIT Press, 1988.

[20] D. Jurafsky and J. H. Martin, “Speech and Language Processing. An Introduction to Natural Language Processing”, Computational Linguistics and Speech Recognition. second. Prentice Hall, 2008

[21] D. Surendran, and G.-A. Levow, “Dialog act tagging with support vector machines and hidden Markov models.” In: INTERSPEECH-2006, pp. 1950–1953, 2006.

[22] D. Verbree, R. Rienks, and D. Heylen, “Dialogue-act tagging using smart feature selection; results on multiple corpora”, in: 2006 IEEE Spoken Language Technology Workshop. IEEE, pp. 70–73.

[23] E. Cambria and B. White, “Jumping NLP Curves: A Review of Natural Language Processing Research”, in: IEEE Computational Intelligence Magazine 9.2, pp. 48–57, 2014.

[24] E. Costantini, S. Burger, F. Pianesi, “NESPOLE!'s Multilingual and Multimodal Corpus”, in: LREC, European Language Resources Association, 2002.

[25] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, “Learning word vectors for 157 languages”, arXiv preprint arXiv:1802.06893, 2018.

[26] E. Ivanovic, “Dialogue Act Tagging for Instant Messaging Chat Sessions”, in: Proceedings of the ACL Student Research Workshop, Ann Arbor, Michigan, 79–84, 2005.

[27] E. Manishina, B. Jabaian, S. Huet, and F. Lefèvr, “Automatic corpus extension for data-driven natural language generation”, in: Proceedings of the Language Resources and Evaluation Conference (LREC), 2016.

[28] F. Mairesse, M. Gaši´c, F. Jurcíˇ cek, S. Keizer, B. Thomson, K. Yu, and S. Young, “Phrase-based statistical language generation using graphical models and active learning”, in: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 1552–1561. Association for Computational Linguistics, 2010.

[29] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval”, in: Information processing & management 24.5, pp. 513–523, 1988.

[30] H. Janet, “An Introduction to Sociolinguistics”. London and New York: Longman Group Limited, 2013.

[31] H. Kumar, A. Agarwal, R. Dasgupta, S. Joshi, and A. Kumar, “Dialogue act sequence labeling using hierarchical encoder with crf,”, in: AAAI, 2018.

[32] H. P. Luhn, “A statistical approach to mechanized encoding and searching of literary information”, in: IBM Journal of research and development 1.4, pp. 309–317, 1957.

[33] J. Ang, Y. Liu, and E. Shriberg, “Automatic Dialog Act Segmentation and Classification in Multiparty Meetings”, in: Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Vol. 1. IEEE, pp. 1061–1064, 2005.

[34] J. Carletta et al.. “The Reliability of a Dialogue Structure Coding Scheme”, in: Computational Linguistics 23.1, pp. 13–31, 1997.

[35] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, M. Kronenthal, G. Lathoud, M. Lincoln, A. Lisowska, I. McCowan, W. Post, D. Reidsma, P. Wellner, “The AMI Meeting Corpus: A Pre-announcement, in: Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction”, MLMI'05, Springer-Verlag, Berlin, Heidelberg, 2006, pp. 28-39.

[36] J. Gang and J. Bilmes, “Dialog Act Tagging Using Graphical Models”, in: Proceedings. (ICASSP ’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Vol. 1. IEEE, pp. 33–36, 2005.

[37] J. J. Godfrey, E. C. Holliman, and J. McDaniel, “SWITCHBOARD: telephone speech corpus for research and development”, in: Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on. Vol. 1. IEEE, pp. 517–520.

[38] J. Jia,, “The study of the application of a web-based chatbot system on the teaching of foreign languages”, in: Proceedings of Society for Information Technology & Teacher Education International Conference. Atlanta, GA, USA: Association for the Advancement of Computing in Education (AACE), pp. 1201–1207, 2004.

[39] J. R. Firth, John R, “A synopsis of linguistic theory, 1930-1955”, in: Studies in linguistic analysis, 1957.

[40] J. Weizenbaum, “Eliza—a computer program for the study of natural language communication between man and machine,” Communications of the ACM, vol. 9, no. 1, pp. 36–45, 1966.

[41] J. Y. Lee and F. Dernoncourt, “Sequential short-text classification with recurrent and convolutional neural networks,” in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (San Diego, California), pp. 515–520, Association for Computational Linguistics, June 2016.

[42] J.L. Austin, “How to Do Things with Words”, Oxford University Press, New York, USA, 1962

[43] J.-M. Benedí, E. Lleida, A. Varona, M.-J. Castro, I. Galiano, R. Justo, I. L. D. Letona, A. Miguel, “Design and Acquisition of a Telephone Spontaneous Speech Dialogue Corpus in Spanish: Dihana”, in: 5th LREC, 2006, pp. 1636-1639.

[44] J.R. Searle, “Expression and meaning: Studies in the Theory of Speech Acts”, Cambridge University Press, Cambridge, UK, 1979

[45] J.R. Searle, “Speech Acts: An Essay in the Philosophy of Language”, Cambridge University Press, Cambridge, UK, 1969.

[46] J.R. Searle, P. Cole and J.L. Morgan, "Speech Acts: Syntax and Semantics”, Volume 3, pages 59-82. Academic Press, New York, USA, 1975.

[47] J.R. Searle. and D. Vanderveken, “Foundations of Illocutionary Logic”, Cambridge University Press, Cambridge, UK, 1985.

[48] K. E. Boyer et al., “Dialogue Act Modeling in a Complex Task-oriented Domain”, in: Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL ’10). Tokyo, Japan: Association for Computational Linguistics, pp. 297–305, 2010.

[49] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval”, in: Journal of documentation, 1972.

[50] K. S. Jones, “Document retrieval systems, Taylor Graham Publishing”, London, UK, UK, chapter A Statistical Interpretation of Term Specificity and Its Application in Retrieval, pp. 132-142, 1988.

[51] Kim, “Convolutional neural networks for sentence classification,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), (Doha, Qatar), pp. 1746–1751, Association for Computational Linguistics, Oct. 2014.

[52] L. Chiticariu, Y. Li, and F. R. Reiss, “Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!”, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics, pp. 827–832.

[53] L. Levin, C. Langley, A. Lavie, D. Gates, D. Wallace, K. Peterson, “Domain Specific Speech Acts for Spoken Language Translation”, in: Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue. Sapporo, Japan, 2003.

[54] L. Màrquez et al., “Semantic role labeling: an introduction to the special issue”, in: Computational linguistics 34.2, pp. 145–159, 2008.

[55] M. Baroni, G. Dinu, and G. Kruszewski, “Don’t count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors”, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Baltimore, Maryland: Association for Computational Linguistics, pp. 238–247, June 2014.

[56] M. F. McTear, “Spoken dialogue technology: enabling the conversational user interface”, in: ACM Computing Surveys (CSUR) 34.1, pp. 90–169, 2002.

[57] M. G Core and J. Allen, “Coding dialogs with the DAMSL annotation scheme”, in: AAAI fall symposium on communicative action in humans and machines. AAAI, pp. 28–35, 1997

[58] M. Henderson, B. Thomson, J. D. Williams, “The Second Dialog State Tracking Challenge”, in: Proceedings of SIGDIAL, ACL Association for Computational Linguistics, 2014.

[59] M. Kay, J. Gawron, P. Norvig, Verbmobil, “A Translation System for Faceto-Face Dialog”, CSLI, Stanford, CA, 1994.

[60] M. McTear, Z. Callejas, and D. Griol, “The conversational interface: talking to smart devices”, Cham: Springer, 2016.

[61] M. Mitchell, D. Bohus, and E. Kamar, “Crowdsourcing language generation templates for dialogue systems”, in: Proceedings of the INLG and SIGDIAL 2014 Joint Session.

[62] M. Schuster, and K. K. Paliwal, “Bidirectional recurrent neural networks”. IEEE Transactions on Signal Processing, 45(11), 2673-2681, 1997.

[63] M. Tavafi et al. “Dialogue act recognition in synchronous and asynchronous conversations”, in: Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2013). Vol. 13. Association for Computational Linguistics, pp. 117–121, 2013.

[64] N. Webb and M. Ferguson, “Automatic Extraction of Cue Phrases for Cross-corpus Dialogue Act Classification”, in: Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING ’10). Association for Computational Linguistics, pp. 1310–1317, 2010.

[65] P. M.Nadkarni, L. Ohno-Machado, and W. W. Chapman, “Natural language processing: an introduction”, in: Journal of the American Medical Informatics Association 18.5, pp. 544–551, 2011.

[66] R. Al-Rfou, B. Perozzi, and S. Skiena, “Polyglot: Distributed word representations for multilingual NLP”, arXiv preprint arXiv:1307.1662, 2013.

[67] R. Bellman, and Karreman Mathematics Research Collection, “Adaptive Control Processes: A Guided Tour”, Princeton Legacy Library. Princeton University Press. isbn: 9780691079011, 1961.

[68] R. Fasold and J. Connor-linton, “An Introduction to Language and Linguistics”, Cambridge University Press, 2006.

[69] R. Fernandez, and R. W Picard, “Dialog act classification from prosodic features using support vector machines”, in: Speech Prosody 2002, International Conference, pp. 291–294, 2002.

[70] R. Serafin and B. D. Eugenio, “FLSA: Extending Latent Semantic Analysis with Features for Dialogue Act Classification”, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL ’04). Association for Computational Linguistics, 2004.

[71] S. Elizabeth et al., “The ICSI Meeting Recorder Dialog Act (MRDA) Corpus”, in: Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue. Cambridge, Massachusetts, USA: Association for Computational Linguistics, pp. 97–100, 2004.

[72] S. Grau, E. Sanchis, M. J. Castro, D. Vilar, “Dialogue Act Classification Using a Bayesian Approach”, in: Proceedings of the 9th International Conference Speech and Computer, 495–499, 2004.

[73] S. Hochreiter, and J. Schmidhuber, “Long Short-Term Memory. Neural Computation”, 9(8), 1735-1780, 1997.

[74] S. Keizer, “Dialogue Act Classication: Experiments with the SCHISMA Corpus”, Tech. rep., University of Twente, October 2002.

[75] S. Larsson, and D. R. Traum, “Information state and dialogue management in the TRINDI dialogue move engine toolkit”, in: Natural language engineering 6.3&4, pp. 323–340, 2000.

[76] S. N. Kim, L. Cavedon, and T. Baldwin, “Classifying Dialogue Acts in One-on-one Live Chats”, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP’10). Cambridge, Massachusetts: Association for Computational Linguistics, pp. 862–871.

[77] T. Andernach, M. Poel, and E. Salomons, “Finding classes of dialogue utterances with kohonen networks”, in: Workshop Notes of the ECML / MLnet Workshop on Empirical Learning of Natural Language Processing Tasks, pp. 85–94, 1997.

[78] T. H. Bui, M. Rajman, and M. Melichar, “Rapid dialogue prototyping methodology,” in: International Conference on Text, Speech and Dialogue. Springer, 2004, pp. 579–586.

[79] T. Klüwer, H. Uszkoreit, and F. Xu, “Using Syntactic and Semantic Based Relations for Dialogue Act Recognition”, in: Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING ’10). Beijing, China: Association for Computational Linguistics, pp. 570–578, 2010.

[80] T. Mikolov, G. Corrado, K. Chen, and J. Dean, “Efficient estimation of word representations in vector space,” pp. 1–12, 01 2013.

[81] T. Mikolov, Q. V. Le, and I. Sutskever, “Exploiting Similarities among Languages for Machine Translation, in: CoRR abs/1309.4168. arXiv: 1309.4168, 2013.

[82] T. Wu et al., “PostingAct Tagging Using Transformation-Based Learning”, in: Foundations of Data Mining and knowledge Discovery. Ed. By Tsau Young Lin et al. Springer, pp. 319–331, 2005.

[83] V. Fromkin, R. Rodman, and N. Hyams, “An introduction to language”, tenth. Cengage Learning, 2013.

[84] V. K. R. Sridhar, S. Bangalore, and S. Narayanan, “Combining lexical, syntactic and prosodic cues for improved online dialog act tagging”, in: Computer Speech & Language 23.4, pp. 407–422, 2009.

[85] V. Rus et al., “Automated Discovery of Speech Act Categories in Educational Games.” In: International Educational Data Mining Society, 2012

[86] https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[87] https://keras.io/

[88] https://www.tensorflow.org/

[89] ဒေါက်တာအောင်မြင့်ဦး၊ လူမှုဘာသာဗေဒ သဘောတရား

[90] ဒေါက်တာအောင်မြင့်ဦး၊ လူမှုဘာသာဗေဒ မိတ်ဆက်

[91] ဒေါက်တာခင်အေး၊ အတ္ထဗေဒနိဒါန်း

[92] မောင်ခင်မင်(ဓနုဖြူ)၊ လက်တွေ့ အတ္ထဗေဒနိဒါန်း

[93] မောင်ခင်မင်(ဓနုဖြူ)၊ အတ္ထဗေဒနိဒါန်း

[94] မဥမ္မာစိုး၊ မြန်မာဘာသာစကားရှိ လမ်းညွှန်မှု အသုံးများ၊ ပါရဂူဘွဲ့အတွက် တင်သွင်းသောကျမ်း၊ မြန်မာစာဌာန၊ ရန်ကုန်တက္ကသိုလ်

[95] မမြင့်မြင့်စန်း၊ မြန်မာဘာသာစကကားတွင် ဘာသာစကားအသုံးနှင့် အနက်အဓိပ္ပာယ်ဆက်သွယ်မှု၊ ပါရဂူဘွဲ့အတွက် တင်သွင်းသော ကျမ်း၊ မြန်မာစာဌာန၊ ရန်ကုန်တက္ကသိုလ်

APPENDICES

Appendix A: The 42 clustered SWBD-DAMSL Dialogue Act Tags

Dialogue Act	Description
3rd-party-talk	Statements made by anyone other than the main speakers.
Acknowledge (backchannel)	Acknowledgements like "Okay.", "Uh-huh.", or "Yeah." indicating that the other person's words were heard.
Action-directive	Commands or prompts for action such as "Go ahead." or "You've got to...".
Affirmative non-yes answer	Affirmative descriptive or narrative statements answering a question.
Agree/accept	Agreement signals like "Right.", "Okay.", or "Yeah." indicating acceptance of the previous speaker's words.
Apology	Apologies such as "I'm sorry." or "Excuse me.".
Appreciation	Expressions of gratitude or appreciation like "That's good." or "Oh wow.".
Backchannel in question form	Rhetorical questions prompting more information, e.g., "Oh, really?" or "Is that right?".
Collaborative completion	Completing the other person's statement.
Conventional-closing	Closing statements like "Bye." and "Nice talking to you.".
Conventional-opening	Opening statements such as "Hello.", "My name is..." and "How are you doing?".
Declarative wh-question	Wh-questions in declarative form.
Declarative yes-no-question	Yes-No questions in declarative form.
Dispreferred answer	Utterances expressing partial disagreement with the other person.
Downplayer	Responses that downplay sympathy or compliments like "That's all right." or "It happens.".
Hedge	Softening statements to prevent commitment, e.g., "I don't

	know.", "Maybe..." or "I'm not sure.".
Hold before answer or agreement	Fillers preceding an answer like "Uhm..." and "Let's see...".
Maybe/accept-part	Expressions of doubt about accepting what was previously said, such as "Maybe.", "I guess." and "I don't know.".
Negative non-no answer	Descriptive or narrative statements giving a negative answer to a question.
No answer	Negative responses like "No." or "Huh-uh.".
Non-verbal	Non-verbal utterances like laughter and coughing.
Offers, options and commits	Offers, commitments, or options presented by the speaker, e.g., "Let me...", "We could...", "I'll do it.".
Open-question	Open-ended questions like "What do you think?" or "How about you?".
Or-clause	Yes-No questions with an additional choice, e.g., "Or is it more of a company?".
Other	Utterances that don't fit other categories, often as "Okay.".
Other answer	Answers that don't fit other categories.
Quotation	Repetitions of someone else's words.
Reject	Rejections of previous statements, typically "No.".
Repeat-phrase	Utterances repeating part of the previous one.
Response acknowledgement	Acknowledgements of a previous answer, e.g., "Oh, okay." or "Oh, I see.".
Rhetorical-question	Rhetorical questions meant to make a point, e.g., "Can't you do anything right?".
Self-talk	Self-directed speech such as "Oh, what was it?".
Signal-non-understanding	Indications of misunderstanding like "Huh?", "Pardon me?" or "What do you mean?".
Statement-non-opinion	Undisputable descriptive or narrative statements.
Statement-opinion	Disputable viewpoints, from personal opinions to general facts.
Summarize/reformulate	Summarizations or reformulations of another's words.
Tag-question	Tag questions like "Okay?", "Right?", "Don't they?" and "Isn't it?".

Thanking	Expressions of gratitude like "Thanks".
Uninterpretable	Incomplete utterances that can’t be interpreted.
Wh-question	Wh-questions starting with words like "who", "what", or "how".
Yes answer	Positive responses like "Yes." or "Yeah.".
Yes-no-question	Questions that can be answered with "yes" or "no".