CHAPTER 1
INTRODUCTION

1.1. Introduction to English-Myanmar Bidirectional Translation Chatbot

Chatbot (also known as Conversational agent) is a program that generates response based on given input to emulate human conversations in text or voice mode. These applications are designed to simulate human-human interactions. Chatbots are predominantly used in business and corporate organizations including government, non-profit and private ones. Their functioning can range from customer service, product suggestion, product inquiry to personal assistant. Many of these chat agents are built using rule based techniques, retrieval techniques or simple machine learning algorithms. In retrieval based techniques, chat agents scan for keywords within the input phrase and retrieves relevant answers based the query string. They rely on keyword similarity and retrieved text is pulled from internal or external data sources including World Wide Web or organizational database. Some other advanced chatbots are developed with Natural Language Processing (NLP) techniques and machine learning algorithms. Also, there are many commercial chat engines available, which help build chatbots based on client data input. Recently there have been major increased or interested in use and deployment of dialogue generation systems. Many major technology companies are using virtual assistant or chat agent to fill the needs of customers. Some of them include Google's Google Assistant, Microsoft's Cortana and Amazon's Alexa. Though they are primarily question answering systems, their adoption by major corporations has peaked interest in customers and seems promising for more advanced conversational agent system research and development.

This system is a chatbot that receives users input which can be in Myanmar or English language and translate the input sentence to the respective languages. This system is implemented in Facebook Messenger. It is like exchanging messages with someone through chat to them to ask for a translation. They are intended to make it easy to use a translation platform. They simulate conversations with humans, to make the use of a translation service somewhat engaging or less impersonal. This chatbot converts the input sentence to its respective language according to input language using

Statistical Machine Translation (SMT) which is based on statistical models trained on bilingual and monolingual corpora. Basically, two probabilistic models are used, a translation model which is trained on bilingual corpora and a language model which is trained on monolingual corpora. SMT has many advantages, it is language independent, easy, cheap and fast to build. Many tools for training and decoding are freely available now. Also the huge bilingual and monolingual corpora needed for training are available for many language pairs. The current state of the art in SMT is the Phrase-based Statistical Machine Translation (PBSMT) because it uses longer translation units than the initial word-based models. By these means, more contextual information is captured by the translation model, which improves the translation quality.

1.2. Aim and Objectives of the Thesis

The main aim of this thesis is to analyze phrase-based statistical machine translation performance between English and Myanmar language pair. The following are the objectives of the thesis.

- to develop a translation chatbot and implement in Facebook Messenger
- to make easy and convenient to users
- to build a translation platform that can be accessed easily
- to build a system that can provide students in learning English language easily

1.3. Scope of the Thesis

This proposed system receives the input as text message. This system can support only English and Myanmar language, so the input sentence must be English or Myanmar sentence. Phrase-based SMT translates phrases as atomic units, so the input sentence is segmented into phrases and each phrase is translated into target language and then translated phrases are reordered. Phrase-based SMT uses distance-based reordering which is the local reordering. The local reordering is not good enough for the different order language pairs. English and Myanmar languages are very different in reordering. Therefore this system can translate only the short sentence well. This translation system is trained with 10K Myanmar sentences of the ASEAN-MT Parallel Corpus which is a parallel corpus in the travel domain. Therefore, this proposed chatbot can translate the sentence which is in the travel domain appropriately. For the sentences

which is out-of-domain such as medical domain, political domain and so on, this system cannot translate appropriately.

1.4. System Overview

In this proposed system, the input text sentences from the user via Messenger are acquired from the Facebook Messenger API. The Facebook server sends webhooks with request message to the URL of the ngrok web server, which redirects the request to the localhost where the messaging app and all the required library files for the Translation Model are hosted. The sentence received sent with the webhook by ngrok server is checked whether it is English sentence or Myanmar sentence. If the user's input is the English sentence, the sentence is parsed to the English-to-Myanmar PBSMT Model or if it is the Myanmar sentence, the Myanmar-to-English PBSMT Model is used. PBSMT model is composed of two main components called translation model (TM) which is trained from the bilingual corpus (especially source and target languages) and language model (LM) which is trained on monolingual corpus (especially target language). TM is responsible to map the source language to target language and associated with the adequacy. TM generates the broken sentence after mapping source language to target language. LM is responsible to distinguish good target usage, so it is associated with fluency. The LM takes the broken sentence generated from TM and contributes to select the fluent sentence in target language. The translated output from the model after parsing the request is sent to the Facebook Server via Webhook by the ngrok. Finally the translated sentence is replied to the user by the Facebook Server.

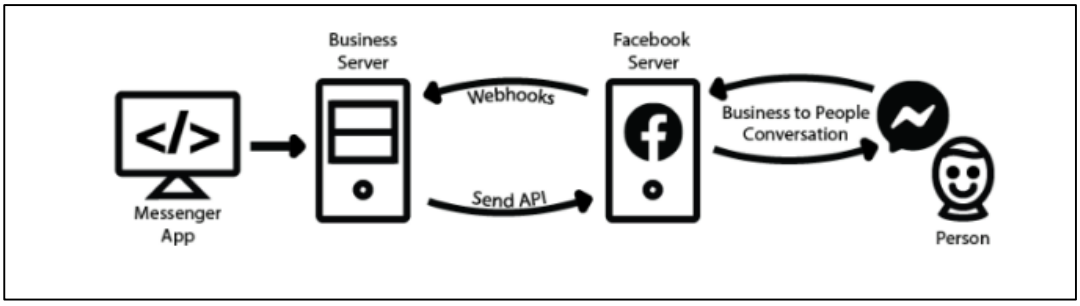

Figure 1.1. Workflows of the Messenger Chatbot

1.5. Outlines of the Thesis

The proposed system is the Myanmar-English Bidirectional Translation chatbot system that is implemented in Facebook Messenger using Statistical Machine

Translation(SMT). This dissertation is composed of five chapters and references, list of tables and list of figures. In Chapter 1, an introduction on chatbot and machine translation, objectives of the system, scope of the thesis, system overview and outline of the thesis are described. In Chapter 2, literature review concerned with chatbot and machine translation are presented. In Chapter 3, the background theory concerned with the statistical machine translation is explained in details. In Chapter 4, the design and implementation of the proposed system is presented. In Chapter 5, the discussion, conclusion, limitation and further extension are described.