

# 1. INTRODUCTION

Humans are individual and social beings that interact with each other. In interacting, language is required to convey purpose and objectives to others. Without language, it can be difficult to convey those purposes and objectives. The function of language for human beings is not only as an intermediary to express themselves, feelings, thoughts, desires, and needs, both as individual and social beings, language is also a tool of integration and social adaptation between humans in developing their civilization [1]. Therefore, language is an important component in everyday life that becomes a means of interacting and communicating between individuals.

Indonesia is one of the countries rich in regional language diversity. The number of regional languages in Indonesia is very large. This is because each province has several regional languages at once. Of the 34 provinces in Indonesia, 718 regional languages have been identified as of 2020. One of the regional languages with the most speakers is Javanese. The Javanese language is a communication media that is primarily spoken by ethnic Javanese people in the central and eastern parts of Java Island. The Javanese language has three levels, namely ngoko, madya, and krama. In everyday life, Javanese ngoko is the Javanese language most commonly used by ethnic Javanese people. The number of regional languages in Indonesia makes not all Indonesians master it. Most Indonesians only speak the regional language used in their region.

The ability to master the language will certainly make it easier to interact and communicate. But some individuals only use certain languages. Machine translation comes as one of the solutions to solve language translation problems. Machine translation is an automatic machine that can process language transfer from one language to another. Machine translation was created to facilitate communication between individuals of different languages.

Machine translation has several approach methods, statistical-based approaches, example-based approaches, and rule-based approaches. Machine translation with statistical approaches is considered capable of overcoming deficiencies from previous approaches, such as rule-based approaches and example-based approaches [2]. Also, in terms of accuracy, machine translation with statistical approaches has better accuracy when compared to machine translation with a rule-based approach [2].

In this research, a phrase-based statistical approach was used in the machine translation Javanese-Indonesian. Statistical machine translation is a type of machine translation where the result is produced based on a statistical model whose parameters are taken from the analysis of parallel corpus [3]. The main idea of the statistical machine translation is that each sentence in the target language is the result of a translation of a sentence from the source language with a certain probability. The translation result is a sentence that has the highest probability. In the construction of the statistical machine translation, the main data source needed is the corpus. Corpus consists of two types, that is parallel corpus and monolingual corpus.

Corpus is a collection of texts both spoken and written in print or electronic media that can be used as a data source. To build a statistical machine translation, parallel corpus containing copies of the source language text and the target language and monolingual corpus containing only text in one language is the target language. Parallel corpus serves to form a translation model, while monolingual corpus serves to form a language model.

The research of regional language statistical machine translation has previously been done by several researchers. Some of these studies include Improving The Accuracy of Statistical Machine Translation by Increasing The Quantity of Monolingual Corpus [4] [5] [6]. The Effect of Parallel Corpus Enhancement on Statistical Machine Translation Accuracy [7] [8]. Indonesian And Javanese Machine Translation Using Phrase-Based Statistical Methods [2] [9]. Statistical Machine Translation In Lampung Dialect Api To Indonesian [10]. Two-Way Translator Application Web-Based Indonesian-Sambas Malay Language Using Moses Decoder [11]. Use of Pivot Language on Statistical Machine Translation English to Sambas Malay Language [12]. Algorithm for Sharing Phrases In Sentences To Improve Accuracy of Statistical Machine Translation Indonesian-Bugis Wajo Language [13], and Statistical Machine Translation Accuracy Test (MPS) Indonesian to Sambas Malay And Statistical Machine Translation (Mps) Sambas Malay to Indonesian [1].

The purpose of this research to produce a statistical machine translation that can be implemented to perform Javanese-Indonesian translation and to know the influence of the main data sources of statistical machine translation namely parallel corpus and monolingual corpus on the quality of statistical machine translation Javanese-Indonesian.