Chapter 4

Evaluation of Machine Translation Output

4.1. Introduction

Evaluation has always been a major concept for researchers of MT. It is axiomatic that evaluation has proven complicated, traumatic, and misleading; yet both revelatory and useful. Its importance has been highlighted since the ALPAC report in 1966 aforementioned in Chapter Two of the current research. ALPAC had nine recommendations for future US research in translation, three of which directly recommended further work on evaluation. Such recommendations resonate nowadays and MT evaluation is important now as it used to be.

To this end, White (2003) believes that MT evaluation is significant for the fact that MT costs great deal of money to research, design, implement and complete a system with knowledge germane to the subject areas which the system is to translate. As a result, one needs to know whether the investment is worth making.

Although indispensible to the field, it should be noted that evaluating MT is a thorny task. This is due to the fact that evaluating translation is intrinsically hard. There can probably be more than one way to translate the same document, and translators may disagree on the best way to translate a document. As pointed out by Bowker (2000: 183), "the primary difficulty surrounding the issue of translation evaluation is its subjective nature – the notion of quality has very fuzzy and shifting boundaries". The situation is aggravated in the case of MT and particularly SMT as Cancedda et al., (2009:2) postulate that "machine learning techniques typically rely on some kind of cost optimization in order to learn relationships