Evaluation of Machine Translation Systems:
The Translation Quality of Three Arabic Systems

Name of researcher: Yasmin Hikmet Abdul Hamid Hannouna
Subject/major: Linguistics & Translation
University name, department name: Dept. of Translation, College of Arts, Al-Mustansiriyyah University, Iraq
Year of award: 2004

Abstract:

Evaluation is an implicit aspect of all human activity. With respect to MT, it remains an open fundamental issue and one of the most important stages in the life cycle of an MT system. The present evaluation study investigates the overall quality of three currently available English-into-Arabic MT systems. The evaluation deals with selected quality characteristics and various text types. This is to bear on their ability and extent to satisfy specific requirements and help users to fulfill their tasks. The theoretical construct adopted in this study is based on the Framework for Evaluation of MT in ISLE (FEMTI, 2003) which is the most recent and comprehensive model of evaluation. The evaluation, constitutes a standard test bed application of this methodology (i.e., task-oriented testing and benchmark testing). The proposed model for the functional criteria is a black-box type, comparative and adequacy-oriented evaluation. As for the non-functional criteria, the evaluation model is said to be the comparative performance and adequacy-oriented type. The sample represents a total of 268 English sentences taken from twelve various special-domain texts. Some computational criteria have also been evaluated. These systems have been tested under experimental conditions by two evaluators. Detailed analyses and classification of the results concerning the selected criteria are presented with Excel tables, charts and graphs. The overall comparison of the three systems in terms of quality assessment of both criteria and texts level confirm that English-into-Arabic MT systems suffer from serious drawbacks especially related to the grammar and meanings of the translated sentence. Their output reflects many deficiencies in translating various text types and they all need serious improvements. In addition, the end user can use these systems to grasp the general idea of the source text, or translate short and simple texts. Three major types of problems have been identified with these systems: a.) cognitive, b.) linguistic and c.) operational. The operational problems are attributed to certain impediments in measuring the speed of translation and some limitations relevant to the design and performance of the user dictionaries of these systems. Having identified these problems, the researcher then investigates their possible sources. On the basis of these findings, a number of suggestions and recommendations are made.

Key words: Machine translation, evaluation of MT systems, black-box evaluation, task-oriented testing, benchmark testing, functional criteria, non-functional / computational criteria, operational problems.