

4.3 Human vs. Automatic Evaluation of MT

MT evaluation could be performed by either human or machine depending on which type is being adopted as mentioned in 4.2. Humans usually rely on natural language and world knowledge, and their intervention can therefore be useful since their judgment of quality tends to be more accurate than any automatic measure (Cancedda et al., 2009). Moreover, Cancedda et al., favor fully automated evaluations on the types 4.2.2. (b) and 4.2.3. As the current assessment falls under type 4.2.1, human evaluation is the optimal alternative.

Nevertheless, human judgments tend to be subjective and have been shown to vary considerably from one assessor to another. As-Safi (2011:13) points out “An objective assessment of legal texts can be achieved contrary to literary texts, as literature is inherently subjective and more than one version or rendition are expected due to translation indeterminacy.”

White (2003:214) looks at the notion of ‘subjectivity’ from a different angle where he admits that while it is difficult to get around subjectivity, we could perhaps take advantage of it. To this end, he states:

despite the disagreement we are likely to have about translation correctness, we still strongly agree about linguistic intuitions in everyday life. We can talk to each other, read works that are hundreds of years old, order food, and so on, with very high confidence that, despite likely differences in our cultural or cognitive models of reality, we fully understand and agree about the meaning of the expressions and the event as a whole. Could we not capture these linguistic intuitions as means of measuring MT?

Since the current research adopts human evaluation, it will neglect the different methods of automatic evaluation and focus on the different cases of human evaluation. Such selection does not by any means underestimate automatic evaluation as it has become an active research topic. It is important to notice here