between the input and output data". We can conclude that such techniques are relevant to machine and are very different to the ones adopted by humans.

White (2003) stresses the difficulty of MT evaluation which measures a certain attribute of something against a standard for that attribute. For this to happen, there need to be an identifiable "correct" or "best ideal" against which to compare the relevant attribute of the individual item being measured. The most obvious standard for MT, i.e. the "right" translation, is the very thing translation itself cannot provide. (ibid: 213)

4.2 Types of MT Evaluation

Cancedda et al., (2009) suggest three different types of evaluation which require a quantitative measure of quality:

4.2.1 Assessing whether the output of an MT system can be useful for a specific application (absolute evaluation);

4.2.2 (a) Comparing systems with one another, or similarly (b) assessing the impact of changes inside a system(relative evaluation);

4.2.3 In the case of systems based on learning, providing a loss function to guide parameter tuning;

White (2003:223-232), on the other hand, proposes a descriptive model of evaluation types. He organizes the different types into the following categories:

Feasibility tests: an evaluation of the possibility for a particular feat to be accomplished at all, or for a particular approach, whether it has any actual potential for success after further research and implementation

Internal evaluation: unlike feasibility testing which is designed to show the potential of an approach, internal evaluations show that implementation of such an