that automatic evaluation cannot only benefit MT but also human translation evaluation.

With respect to human evaluation, White (2003) proposes three case scenarios which can serve as an evaluation method to be adopted by MT researchers:

Case 1 (output only): This is the simplest way in which an MT developer scrutinizes an output and indicates whether it is of good language. In this case, the following metric is used:

Look at each sentence, one at a time; **EITHER**: the sentence is completely good English; **OR**: the sentence is degraded by up to ***n*** errors. **OTHERWISE** the sentence is wrong

Figure 1.Case 1: counting errors.

This metric is used to score output sentences expressing either a quantitative measure (by sentence, document, or whole test set), a qualitative method (by characterizing the errors in some way), or both. The shortcoming of such case is that we know nothing about the input and therefore our characterization may not really help to improve the system as translation issues are completely neglected.

Case 2 (input and output): Here an MT developer looks at both the input and output, but the metric has to be changed as there are now two parameters: whether the output is fluent (intelligibility), and whether the information in the input is conveyed in the output (fidelity).