Rewrite of Discussion Section

Katharina Probst

January 10, 2003

1 Controlled Elicitation or Learning from Uncontrolled Corpora?

In our work, we are interested in what we call a 'resource-rich/resource-poor situation'. In other words, NLP resources such as a parser and a grammar are available for one of the translation languages, whereas they are not necessarily obtainable for the other language. If we classify learning situations by this criterion, there is still the question of whether a large uncontrolled corpus is available. Let us assume for sake of argument that this is not the case. We will emphasize the strenghts of our approach in this, assumed, situation. It will then become clear that even if a large uncontrolled corpus is available, our approach can be a valid contribution to learning from the large corpus.

Note: where does the following go?? Two points should be made clear from the start: first, our learning approach is applicable equally to uncontrolled and controlled corpora. However, as we will argue, learning is facilitated if a small but controlled corpus is used as training data. Second, the rules that our system learns can be used as a stand-alone translation grammar (where the grammar is possibly improved by hand or by the translation correction tool). In addition, the grammar can also be integrated with other MT systems, such as an EBMT system, to function in a multi-engine system.