[*MT evaluation: basis for future directions*. Proceedings of a workshop...2-3 November 1992, San Diego]

Evaluation of Lexical Coverage

Gudrun Magnusdottir
Department of Swedish and Computational Linguistics
University of Gothenburg

The average number of lexical units in marketed machine translation systems is about 20,000 items of basic vocabulary. The number of items, the composition of the vocabulary, and how the dictionary is stored are crucial to achieving quality and economy with a machine translation system.

In a glass-box evaluation it is easy to list the lexical entries and compare them against a frequency list of the client's documents. Closed-class lexical entries and verbs need to be comparatively well covered in the lexicon that is delivered with the system, whereas nouns and adjectives should be mainly adapted to and updated from the client's texts.

The cost of lexical adaptation can be easily calculated using the number of lexical entries that need to be entered and the time it takes to enter them. A user enters a sample list of nouns, verbs, and adjectives and then averages the time (T) it takes to enter each category (CAT). This figure is used as the basis for calculating the cost (T(CAT) x salary) for the complete adaptation process. The result will be higher than the actual cost, since the user will gain increased speed as he/she becomes more experienced.

The number of entries that need to be entered is derived from a comparison between the client's texts and the system's dictionary. The entries are organized into categories, and the calculation is based on the number (N) of entries for each category. The is such a notable difference in the time required to enter nouns, adjectives, and verbs that each category needs to be calculated separately.

Thus the cost for entering each category of words is:

(T(CAT) x salary x N = cost per category

In a black-box evaluation, the documents need to be reviewed for words not present in the lexicon. The procedure in itself is simple, and the results should be analyzed according to the same simple algorithm in order to estimate the cost.

The sum of the cost for each of the three categories will represent the approximate expense entailed in introducing a machine translation system into the company's environment. The system will then perform at its full level of competence, but that competence will not have been tested by the company.

Testing the competence of a system is hardly possible for the average purchaser. To do so requires technical knowledge of natural language processing, linguistics, computer science, and languages. However, the purchaser can study translation quality and see in what ways the system, even though it has complete lexical coverage, still makes mistakes. The most common problems are inadequate translations of idiom-like constructions, pronouns, and adjectives. If the purchaser is sure that the texts the company needs to translate are relatively free of such constructions, state-of-the-art machine translation will perform nicely.

It is regrettable that machine translation is marketed as a black box when the linguistic data in the system is so crucially important to system performance. In fact, purchasers should not accept black-box evaluations of machine translation systems and should insist that developers be more open about the content of their systems before buying them. This may partly be achieved by less linguistic snobbery, coupled with the willingness to recognize that what works best is not always theoretically clean.