For the Europe of 23 languages: the computer learns to translate

The new European project EuroMatrix develops automatic translation systems for all official languages of the EU.

The number and scope of translations in business, administration, politics and culture is increasing yearly. The principle clients for translations are export-oriented economic sectors, international organisations including the European Union and the United Nations, as well as multinational corporations. The exploding demand for text translations is one of the direct consequences of open trade practices, globalisation, European integration and international cooperation. The legion of translators is growing. However, with the general trend towards specialisation, the demand on the competence of specialised translators also increases. This leads to a general increase in the cost of a good quality translation. The European Commission cannot, for example, send technical and legal translations among the languages of the European Union to the Philippines, India or other low-wage countries.

Translators working for large companies and organisations naturally have been using specialised computer programs for years, which are able to fish appropriate suggestions for translations from the body of previous translations. However, this method cannot render a sentence that is not available in the program's memory.

This is made possible only by automatic translation, which is often represented as being the only realistic solution to managing the communication problems of a multilingual world society. Automatic translation is increasingly viewed by experts as a critical precondition for the enduring preservation of the manifold languages and cultures of our planet.

Although experts agree among themselves on the immense demand for automatic translation, the opinions regarding the possibilities of such systems with regard to complex texts differ widely. While many experts still doubt that machines can deliver adequate translations, there are corporations which are already reducing costs through the use of computers.

The truth of the matter lies not quite in the middle, although both sides are right in their own way. Even outdated and low-quality translation systems, which are currently only of very limited use in business and administration, bring costs down. The real breakthrough in automatic translation, however, still lies before us. And this at a point at which the business sector has more or less lost hope in automatic translation.

For decades, unrealistic expectations were consistently disappointed. From 1975 to 1995, the state of research was largely stagnant. Since 1995, a noteworthy new trend has emerged, breathing new life into the current research: the development of a statistical translation method. Rather than using complex handmade systems of rules requiring decades of concentrated work to refine, one can instead use the body of translated work through the application of automatic learning processes. The result is computer programs which have taught themselves to translate.

The human translator must have perfect command of the target language because it is more difficult to compose stylistically excellent language than it is to simply understand content. The statistical translation programs can be similarly perfected in the language into which they translate. This is done via automatic learning, for which large supplies of monolingual text, found in abundance in the internet, are necessary. These autodidacts of translation systems are still far from perfect, but they are improving year by year.

Technological breakthroughs in automatic translation are to be the aim of a new international research program which is being financed under the 6th framework program of the European Union. New combinations of the best currently available methods for rule-based and statistical translation are to be tested. Additionally, the question of how statistical translation systems can be improved by linguistic methods will be investigated. The ambitious aim is to develop and test translation systems for the 23 official languages of the European Union. An important part of the project is also the organization of competitions among the best translation systems, to which all research centers and firms from Europe and other parts of the world are to be invited to participate. It will naturally not be possible to focus equally on each language pair, but a practical result of the project is to be a continuously updated appraisal of the technological state for each language. This overview in the form of a large tabular matrix has given the project its name.

The project is coordinated by Saarland University represented by Professor Hans Uszkoreit. Dr. Philipp Koehn from the University of Edinburgh is serving as technical/scientific coordinator of the project. Other partners include internationally renowned research groups at Charles University in Prague lead by Prof. Jan Hajic and the Center CELCT in Trento under the direction of Prof. Amedeo Cappelli.

The partnership with representatives of industry is also important: the German firm GroupWare AG is bringing their LOGOS system into the project, LOGOS being one of the largest and most important commercial translation systems, which was made available the year before last in an Open-Source version. The firm Morphologic in Budapest has developed digital dictionaries for many European languages and has its own translation technology.

The project has recently kicked off and is planned for 30 months at a cost of 2.5 million Euros.