EAGLES: Stretching its Wings

For NLP to become commercially viable, standards are needed. A group of European researchers has banded together to get the ball rolling in the right direction.

As the various CEC-funded Linguistic Resource and Engineering (LRE) projects get off the ground, various teams have begun publicizing their activities. Lately, we have been hearing a lot about EAGLES, short for Expert Advisory Group on Language Engineering Standards. This three year LRE project, just off to a start in February, is being coordinated by Consorzio Pisa Ricerche (CPR) under the direction of Antonio Zampolli. Five working groups have been formed, dedicated to text corpora, computational lexicons, grammar formalisms, evaluation and assessment, and spoken language, and each has been appointed a host organization and chairperson. The working groups will be soliciting input from more than thirty European companies, institutions, and networks, and in turn recommending common functional specifications and proto-standards for their respective domains.

The idea for EAGLES was hatched several years ago at meetings where representatives of ACQUILEX, MULTILEX, GENELEX and EURO1RA-7 — all CEC and EUREKA-funded proj ects, which broadly speaking have been addressing lexicon-related issues — came to the not so surprising realization that they were all striving for more or less the same goal. In hindsight, this may seem obvious, but as the CPR's Nicoletta Calzolari points out, the processes have long overshadowed the lexicon in computational linguistics. This is now changing, with the lack of generic basic technology and large-scale language resources widely acknowledged as the prime obstacle in the construction of commercial applications. This shortcoming is now being addressed but the diversity of lexical formats and functional specifications still results in a duplication of efforts. The impulse to agree on a polytheoretical specification for computational lexicons, largely stimulated by the organizational efforts of the indefatigable Zampolli, has since evolved into EAGLES. If everything goes according to plan, Calzolari and UMIST's John McNaught will be compiling an initial set of wide-ranging EAGLES proposals in early 1995.

Will the more than thirty European companies and organizations participating in EAGLES actually be able to agree on anything? Is there enough common ground in theoretical matters to make concrete agreements possible?" Calzolari, for one, believes that EAGLES can make some progress. In corpus tagging, for example, it should be possible to define parameters for morpho-syntactic information as well as standardize some common syntactic descriptions, such as constituency and dependency structures. In addition, a new class of declarative, logic-based formalisms has emerged — frequently subsumed under labels unification or constraint-based grammar. These have become the leading paradigm in contemporary computational linguistics and are therefore ripe for some form of standardized notation. Concerning the semantic realm in lexicon coding, however, Calzolari warns that it will be more difficult to come to agreement here, at least within the near future, as this is still a contentious area.

It will take more than good intentions to define even provisional standards in this rapidly evolving world, but the motivation for the EAGLES participants to rise to the challenge will surely be enhanced by profound necessity. Academic researchers will never attract commercial partners to help exploit the fruits of their endeavors unless they can demonstrate robust systems with broad linguistic coverage, ripe for deployment. In the commercial world, there probably isn't a computer company left in the world with the resources to autonomously fund the development of the large scale linguistic resources needed to reach the mythical payoff; collaboration here is a must. Then, there is always the threat of Japan on the not so near horizon to sharpen people's thinking. While the Japanese have not yet had great success in NLP, many companies there are known to be working on it. If Europe doesn't get its language technology act together, there looms the specter of potential Japanese ascendency in this field. Cameras, consumer electronics, cars — there are plenty of historical paradigms to fuel the imagination and hasten proactive cooperation.

EAGLES Secretariat, Fax +39 50 58 90 55, Email eagles@icnucevm.cnuce.cnr.it