

Another Boost from Luxembourg

The European Commission would keenly like to see European industry take a leading role in language engineering. With this in mind, it has funded another fifteen projects within the framework of the LRE program. Nowadays, bigger is not necessarily better; small – but not too small – is beautiful.

The second round of projects in the CEC's Linguistic Research and Engineering (LRE) program have been approved, and, as you read this now, DG XIII/E's LRE officials are putting the final touches on project synopses in readiness of official approval and subsequent public distribution. The LRE, you will recall, is the successor to the Eurotra program, and as such is the most prominent, but by no means the only Community program, that funds language processing R&D and related activities.

A glimpse of the project synopses reveals that the second round of LRE shared-cost projects has firmly consolidated the trend initiated in the first round towards smaller, more focused efforts at developing useful linguistic tools based on existing technology and cultivating common resources. Whereas in the heyday of Eurotra in the 1980s, the financial sustenance of the twelve national groups was more or less a given from year to year; the LRE program has evolved into a vigorous competition among a large number of consortia for a rather limited amount of money. LRE officials report that eighty-two eligible proposals were received for the second call that was published in October, 1992, exceeding by sevenfold the available funds.

The fifteen projects approved in the second round are a varied lot. Basic "generic" research, while less prominent than in the past, is present in the form of FRACAS, a project to develop a common semantic framework. Evaluation receives some much-needed attention in TEMAA, which focuses on writing tools, TSNLP, which will define tests suites for NLP systems, and SQALE, which will build upon the experiences of several European speech groups in ARPA evaluations.

In the application prototype arena, ANTHEM will develop a system for translating medical diagnoses, COMPASS will develop a text comprehension aid based on commercial bilingual dictionaries, and GIST will develop a text generation system for social security documents. SECC will develop a controlled English checker based on Siemens's METAL parsing engine, SIFT will develop a concept-based retrieval system for computer manual texts, RENOS will build multilingual retrieval software for legal texts, and lastly TRANSTERM will create a toolbox for developing and maintaining terminological resources.

In addition, a number of collaborations will address tools and resources. RELATOR will lay the groundwork for a prototype repository for written and spoken linguistic data, rules, and tools, EUROCOCODA will ensure a healthy European dimension to COCODA, a worldwide initiative dedicated to the development and assessment of speech systems and speech databases, and MULTEXT will assemble and tag multilingual corpora and develop corpus exploitation tools.

An important thrust in this round of projects is an explicit user-orientation. Users are participating in several of the

projects to ensure that goals are relevant and that the results are validated. In this context, users range from what we might ordinarily consider to be end-users, such as suppliers to the medical and legal professions, who might be potential users of an application, to application developers, who might be expected to exploit a given technology. Whereas Eurotra is now perceived as having been too technology-driven – it didn't adequately reflect the needs of a genuine group of users – the pendulum seems to be swinging to the opposite extreme in a few cases, with the Commission funding application development in areas in which industry has clearly demonstrated prior feasibility. A case in point is SECC, the above-mentioned project to develop a restricted English checker, something that Boeing, Carnegie Group, Cap Volmac, and Smart Communications have already done. Be that as it may, the Commission should at least now be confident that project leader Siemens Nixdorf Liège will be able to deliver the goods. And indeed the LRE program needs some successes.

Sadly lacking from the LRE II lineup are the much vaunted "SMEs" (Small and Medium-size Enterprises) which Commission dogma proclaims is the future of linguistic engineering in Europe. LRE program coordinator Roberto Cencioni acknowledges this situation, describing it as something of a philosophical dilemma. "Do we fund lots of little groups?" he asks, "or several large clusters?" Part of the problem is that the project funding procedures cannot be easily accommodated by small companies. On the order of three person-months is required for the initial proposal, consortium meetings-and-technical annex trajectory – and that is before any funding materializes. For seasoned grant writers familiar with the bureaucracy, this can be time well spent. Naturally, that is another matter for small companies new to the game.

On a more profound level, there is the underlying concern that SMEs cannot guarantee the long-term continuity required to nurture these infant technologies to adulthood. If staff members come aboard when Community funding materializes, will they be kept on after the project is over? The research world perceives SMEs as being too occupied with day-to-day business (ie, survival) to innovate or think abstractly. For better or worse, blue-chip names like SNI, GSI-ERLI, SITE, Philips, Cap Gemini Innovation, and Cap debis seem like safer bets for industrial partners.

You might wonder what can be expected in terms of results from the LRE program. As yet, it is too early to tell. The projects launched in the first round are now just getting up to speed and can only be adequately evaluated in another year or two. This year's projects will only be getting off the ground in early 1994; these, too, are due to run a two to three year course. Thereafter, it will also take time for the expertise, ideas and resources to manifest themselves overtly. One difference now, however, is that a number of the projects do promise concrete results. Whereas a project to specify the logic for a formalism to be used in further exploration of a methodology for x can quietly recede into oblivion, one of the declared goals of a project like MULTEXT is to deliver working software tools that will be put into the public domain for fellow practitioners to use. Success and failure will be a magnitude easier to measure, making the stakes higher for both the participants of the projects and the Commission.

Ultimately, however, the fruits of Community-funded linguistic research remain impossible to anticipate; these days, you'll find Eurotra and ESPRIT offspring in odd places. Heartening, yes, but it doesn't make the funding process more certain – or less painful.

CEC DG XIIIIE/4, Bâtiment Jean Monnet, Plateau du Kirchberg, L-2920 Luxembourg; Tel: +352 4301 32859, Fax: +352 4301 34999