

Controlled Language Comes of Age

A Report on the Third International Workshop on Controlled Language Applications, 29-30 April 2000, Westin Seattle Hotel, Seattle, Washington



Bob Clark

It has been six years since the First International Workshop on Controlled Language Applications (CLAW96) was held at the Catholic University of Leuven, Belgium, where "researchers and practitioners from around the world came together for the first time to discuss this exciting new field". What we saw and heard at the Third International Workshop on Controlled Language Applications (CLAW2000) in Seattle, Washington at the end of April clearly demonstrated the distance that this heretofore relatively obscure area of development has travelled. For years, development seemed to be restricted to companies like Caterpillar and Boeing and Controlled Language and Controlled English were virtually synonymous and interchangeable. Workshop papers clearly proved that this is no longer the case.

Ably assisted by local organiser Rick Wojcik (Boeing) and the rest of the programme committee, Geert Adriaens (Lernout & Hauspie), Arendse Bernth (IBM Research), Kurt Godden (General Motors), Teruko Mitamura and Eric Nyberg (Carnegie Mellon University) and Remi Zajac (New Mexico University), the workshop was produced by Jeff Allen, directed by Jeff Allen and featured Jeff Allen in the starring role of Workshop Chair. If there was a competition for "Mr Energy, 2000, Jeff would win hands down (I don't know what he's on but I could use some of it!).

The two-day, full plenary workshop cleverly kicked off each day with detailed demonstrations of applications that would be discussed in papers delivered later in the day. The first of these presented the Boeing Simplified English Checker (BSEC). Based on Boeing's internal AECMA (European Association of Aerospace Industries)-compliant system used since 1990, the product is now being released as a commercial product in collaboration with Lernout & Hauspie.

The programme continued with a paper that was to be delivered by Linda Means (General Motors) on *Training for Controlled Language Processes*. In her absence, Kurt Godden filled in for Linda and continued with his own paper, an update of the General Motors ground-breaking CASL project, *The Evolution of CASL Controlled Authoring at General Motors*. Of the remaining papers presented that day one of the most intriguing was Arendse Bernth's *Easy English: Grammar Checking for Non-Native Speakers*. IBM, along with all the other multination-

als, have realised that nearly all authoring is being done in English with a large portion being produced by people whose mother-tongue is not English. IBM's EasyEnglishAnalyzer (EEA) anticipates the errors that non-native English speakers are likely to make. So far, work has been mainly with Japanese and Germanic languages. Oddly enough, this approach was pioneered years ago by Reference Software, the developers of Grammatik. They had produced regional variations of the software to check English written by French and German speakers when WordPerfect took them over. As far as I know, the non-native authoring solution died on the shelf. Corinne Moore's Invited Paper, *Controlled Language at Diebold Inc.*, was a very revealing case study of a major project that went bad. It is infinitely more useful to hear how not to do it, rather than an endless series of success stories. Other papers included Thierry Declerck's *Towards a Theory of Textual Errors* and Jeff Allen's *Toward the Development of a Postediting module for Raw Machine Translation Output: A Controlled Language Perspective*. The day ended up with the first panel discussion, *Efficient Roll-in and Roll-out of Controlled Language Applications*. The panelists were Jeff Allen (ELRA), Uus Knops (LANT), Becky Letterman (Lucent Technologies) and Rick Wojcik (Boeing). One of the advantages of the small attendance, 50-60 participants, was that the panel discussions quickly involved the whole workshop.

After the second day's demonstrations Ursula Reuther (IAI) presented her *Designing a Multi-Purpose CL Application*. In his *Building Multilingual Controlled Language Performance Checkers*, Andrew Bredenkamp (DFKI) gave an overview of the FLAG (Flexible Language and Grammar Checking) project. Providing another clear indication that Controlled Language has come a long way, Bredenkamp explained that "until now, the automatic checking of controlled language or style has only been available to those very few organisations (Boeing, Caterpillar, etc.) who are willing to invest the relatively large sums required to fund bespoke development. The FLAG project aims to develop tools which will allow for the extremely rapid (and therefore economically viable) development of open client-specific language checkers. They should be open at least in the sense that they can be extended by the client, and not just by the developers... The tools developed in FLAG are inherently multilingual, since a language-independent rule for-

malism has been developed for describing errors - in practice demonstrators for German, English and French are under development." Other papers were *Controlling Plural Ambiguities in Attempto Controlled English (ACE)*, Uta Schwertel (University of Zurich) and *A Word Sense Checking Application for Simplified English*, Heather Holmback (Boeing Phantom Works). The second day's panel discussion, *Controlled Language in the Commercial Translation/Localisation Sector* brought all the issues together. With sixty years in the business between them, the panelists, Scott Bennett (Logos Corporation), Geert Adriaens (Lernout & Hauspie), Butch Pfremmer (Alpnet) and Chris Boudreau (Lionbridge), who actually delivered his position paper by telephone from Ottawa, clearly

defined the role of Controlled Language in multilingual content management. As Geert Adriaens said, "Authoring should be viewed as a translation issue - from uncontrolled to controlled language with Controlled Language serving as a terminology management tool. There is a strong case for developing a broad platform integrating all speech and language solutions." Looking at it this way, it should not be too long before we see Translation Memory moving upstream to become Authoring Memory and Controlled Language moving downstream to be applied to the translated target language. After all, if someone has gone to all the trouble to make the source language compliant with certain standards, should we not ensure the same compliance with the target language? ■