

CL2003: the International Conference on Corpus Linguistics

Andrew Roberts
andyr@comp.leeds.ac.uk

May 2003



Corpus Linguistics researchers from Leeds University, alongside Geoffrey Leech (far right), original raison d'être for the Corpus Linguistics conference.

After the initial success of the first Corpus Linguistics conference in 2001 in honour of Geoffrey Leech's 65th birthday (Rayson *et al.*, 2001), Tony McEnery et al. decided to make it a more permanent fixture in the linguistics calendar, and stage the conference every two years. Hence, it was time for the sequel to commence, and as a result saw hundreds of corpus linguists flocking to Lancaster. The peaceful university campus, located out of town, with its country-side atmosphere, was a lovely setting for the event. Luckily, the weather also blessed us with fine sunshine throughout - much to the relief of the many international participants who were expecting the infamous British climate (i.e., wet and cold!)

Around this time two years ago, Geoffrey Sampson wrote an article in ELSNews regarding the future role of ICAME (in particular, their conferences). He remarked

that ICAME's focus on English research, coupled with its restrictions on conference sizes to ensure a friendly atmosphere, means that not only does it miss out on the ever increasing (and exciting) shift towards non-English language research, but also the next generation of researchers, with energy and fresh ideas, but who are not yet established enough to join the ICAME clique. He went on to comment about the success of CL2001 that had taken place shortly before, and predicted success for its future. ICAME's weaknesses are CL's strengths, which is why Geoffrey's prediction was correct.

The conference kicked off with a day of workshops covering development of learner corpora and multilingual corpora, corpus-based approaches to figurative language, and shallow processing of large corpora. The SProLaC workshop issued its own proceedings (Simov and Osenova, 2003); other papers appeared in the main CL2003 proceedings (Archer *et al.*, 2003). The following four days saw papers presented in three parallel sessions. With approximately 95 papers and 30 posters on offer, it would be impractical to go into any real depth about them. Needless to say, the coverage of the field was well represented. New resources and research were being developed for European minority languages and South-East Asian languages. The usual suspects, such as tagging, parsing, disambiguation, grammars and information extraction were well covered. Unsurprisingly, there was also a strong focus on corpus development, annotation and tools. Even if translation studies, exploiting corpora and semantics are added to the list, it is still not exhaustive - which simply illustrates the breadth of the conference. Of course, the fact that each piece of research presented by definition is linked to corpus, means that there was this common thread throughout, and so as varied as the topics were, they never felt disjointed.

Invited speakers were Michael Hoey, Nancy Ide, Susan Hunston, Geoffrey Sampson and Nicoletta Calzolari. All are well known within the field and offered fascinating and well received talks. Probably the most infamous presentation of the conference was by Tony McEnery, covering his studies of the *f*-word within the BNC. It was even heard that Eric Atwell, a speaker in a parallel session, was actually recommending to his audience to go and see Tony's instead as it was "more interesting!"

The hospitality was excellent throughout the conference, especially the catering, which was of particularly high standard. Naturally, coffee breaks and meal times were the main social occasions. None more so than on the third day, where we were all whisked away by coach to the magnificent Ashton Memorial within the beautiful Williamson Park, which sits high above the main town centre. From here, people could enjoy the lovely scenery, most notably the mountains of the Lake District National Park. The memorial building itself wasn't particularly large, and with so many people in it, personal space was quickly becoming a luxury. If there was anybody with whom you weren't acquainted with - you were by the end of the night! Which was a good thing in my opinion.

The conference was very well organised, and was probably as close to optimal as you could get. Whilst there were approximately 200 participants (representing 30 countries), the atmosphere was still very friendly and informal. Also, three does appear to be the magic number in terms of the number of parallel sessions. At times, choosing one of three talks could be a difficult decision, but at least it gives a degree of flexibility. Any greater and I think that people may begin getting frustrated at missing too many

talks, especially when there is more than one presentation of interest at a given time - as is often the case at some of the larger conferences. Therefore, congratulations should go to Tony McEnery, Dawn Archer, Paul Rayson, Andrew Wilson, plus the many other local staff who helped to ensure an enjoyable and interesting conference. We look forward to CL2005!

References

Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery (eds.) 2003. *Proceedings of the Corpus Linguistics 2003 conference*. UCREL technical paper number 16. UCREL, Lancaster University.

Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie and Shereen Khoja (eds.) 2001. *Proceedings of the Corpus Linguistics 2001 conference*. UCREL technical paper number 13. UCREL, Lancaster University.

Kiril Simov and Petya Osenova (eds.) 2003. *Proceedings of the The Workshop on Shallow Processing of Large Corpora (SProLaC 2003)*. UCREL technical paper number 17. UCREL, Lancaster University.