Vol.2010-NL-197 No.4
2010/7/22

【招待講演】

Discourse processing and citation processing for searching the scientific literature

Simone Teufel †

† **Cambridge University, Computer Science Laboratory**

When scientists read an article under time pressure, they use various signals in the article (some visual, some linguistic, some other) to infer where the information currently most important to them is likely to be found. An automation of this process would bring many advantages for summarisation and search. My past research has been on which of these signals might be both useful and automatable. Some signals used are very sophisticated, whereas others (such as the location of a piece of information) are not. The most interesting signals, but also the most difficult ones to recognise, turn out to be the linguistic indicators. What is also remarkable is the role that citations play in the interpretation process: I claim that the relevance of a citation to a search task is often closely connected to the rhetorical function that the citation plays in the article, and that recognition of citation function is technically feasible.

I will report on previous work and some current work in progress with the Tokunaga group, who have previously done work on identifying the beginning and end points of "citation blocks" around citations, i.e.,the textual area around citations which logically belongs to that citation.