

Human and machine excerption of neologisms

JAKOB HALSKOV & PIA JARVAD

One of the key tasks of the Danish Language Council is to monitor, record and document linguistic changes in the Danish language. To assist in this task, a prototype neologism detector called the *Ordtrawler* (Word trawler) is being developed. The main contribution of the article is a thorough evaluation of the *Ordtrawler's* detection performance on actual linguistic data, but the article also provides a historical perspective on the neology tracking activities of the Language Council. Finally, it describes the advantages and disadvantages of human versus semi-automatic detection of neologisms. The prototype processes large amounts of text and extracts candidate neologisms using a combination of simple filters, collocational statistics and so-called neology markers. The manual evaluation indicates that neology markers are good for optimizing system precision, but at the expense of recall. Certain types of noise such as technical terms and semantically transparent compounds remain to be handled, but diachronic frequency profiling of candidate neologisms appears to be a promising filter.