

Utilising heterogeneous language resources for term extraction in specialised domains

Gisle Andersen
NHH Norges Handelshøyskole Bergen

Although an important player in the maritime industries for many centuries, Norway has not until very recently prioritised the systematic development of an official maritime terminology. In my presentation I report on the development of such a national resource and how the project *Maritim ordbok* (Maritime dictionary) has utilised a multitude of language resources for the purpose of creating a national terminology for the maritime sector. This is now an integral part of *Termportalen* (The Terminology Portal), a repository for terminology connected to the CLARIN/CLARINO research infrastructures. I will describe the role of this resource for purposes such as translation, teaching and language standardisation, and I focus especially on the variety of methods for term extraction that have been/are being applied and experimentally tested in the project. These involve different algorithms for identifying term candidates in general and domain-specific corpora including a domain-specific parallel (translation) corpus and a corpus of popular science, as well as reusing relevant sections from various domain-specific word lists and monolingual and bilingual dictionaries.