

**DOCTORAL CANDIDATE:** Arne Skjærholt  
**DEGREE:** Philosophiae Doctor  
**FACULTY:** Faculty of Mathematics and Natural Sciences  
**DEPARTMENT:** Department of Informatics  
**AREA OF EXPERTISE:** Natural Language Processing  
**SUPERVISORS:** Lilja Øvrelid, Stephan Oepen

**DATE OF DISPUTATION:** 4<sup>th</sup> of December 2019

**DISSERTATION TITLE:** *On morpho-syntactic annotation:  
Methodological investigations of efficiency and  
quality*

Creating modern AI systems requires large amounts of data. In his thesis, Arne Skjærholt explores how the process of creating these data can be managed, and in particular how it can be made more efficient and how to ensure that the data are good enough.

The question of efficiency is explored across several different languages and in the context of several different kinds of system. For morphological processing of Latin and syntactic processing of Norwegian, he investigates the usefulness of using previously created data to partially automate the process and compares these two tasks in view of the different results given by very similar inputs. For English syntax, possible approaches to using untrained workers (“crowdsourcing”) are explored. Finally, he develops a general techniques for evaluating the quality of a syntactic dataset.