

# Studies in Language Structure using Deep Learning

Adam Ek,

Centre for Linguistic Theories and Studies in Probability  
Department of Philosophy, Linguistics and Theory of Science  
University of Gothenburg

Friday 8th of September 2023 at 10:15 in Hörsalen J222,  
Humanisten, Renströmsgatan 6, Göteborg

## Abstract

This thesis deals with the discovery, prediction, and utilization of structural patterns in language using deep learning techniques. The thesis is divided into two sections. The first section gives an introduction to the tools used and the structures in language we are interested in. The second part presents five papers addressing the research questions. The first three papers deal with discovering and predicting patterns. In the first paper, we explore methods of composing word embeddings to predict morphological features. The second paper deals with predicting the depths of nested structures. The remaining three papers deal with using structures in language to make semantic predictions. The third paper explores using dependency trees to predict semantic predicate-argument structures using a rule-based system. The fourth paper explores modeling linguistic acceptability using syntactic and semantic labels. The fifth paper deals with exploring how punctuation affects natural language inference.

## Sammanfattning

Denna uppsats handlar om att upptäcka, predicera och att använda strukturella mönster i språk med maskin inlärning. Denna uppsats är indelad i två sektioner. Den första sektionen ger en introduktion till de de maskininlärningsverktyg som vi använder samt introducerar de strukturella mönster i språk vi tittar på. Den andra delen av denna uppsats presenterar fem artikler jag har producerat för att besvara min forskningsfrågor. Den första artikeln handlar om att sätta samman inbäddningar för att skapa ordinbäddningar, med syftet att predicera morfologiska särdrag. Det andra pappret handlar om att predicera inbäddade strukturer. De resterande tre papprena handlar om att använda strukturer i språk för att predicera soråklig betydelse. Det tredje pappret handlar om att använda dependens träd för att predicera predikat-argument strukturer. Det fjärde pappret handlar om att modellera lingvistisk acceptans med hjälp av semantiska och syntaktiska särdrag. Det femte pappret handlar om att undersöka hur punktuering påverkar inferens i naturligt språk.