In linguistics, words are usually considered to be composed of morphemes: units that carry meaning and are not further subdivisible. The task of this thesis is to create an automatic method for segmenting Czech words into morphemes, usable within the network of Czech derivational relations DeriNet.

We created two different methods. The first one finds morpheme boundaries by differentiating words against their derivational parents, and transitively against their whole derivational family. It explicitly models morphophonological alternations and finds the best boundaries using maximum likelihood estimation. At worst, the results are slightly worse than the state of the art method Morfessor FlatCat, and they are significantly better in some settings.

The second method is a neural network made to jointly predict segmentation and derivational parents, trained using the output of the first method and the derivational pairs from DeriNet. Our hypothesis that such joint training would increase the quality of the segmentation over training purely on the segmentation task seems to hold in some cases, but not in other. The neural model performs worse than the first one, possibly due to being trained on data which already contains some errors, multiplying them.