Compounding is a word-formation process wherein several words, roots, or stems are combined to create novel words. It has been observed in many languages, and often stands on the boundary between word formation and syntax. As such, a multilingual perspective on this process can be valuable for several fields of study, namely morphology, syntax, and typology. In this thesis, we focus on Czech, English, German, Dutch, Russian, French, and Spanish.

We first model compounds in terms of the words that they can be traced back to, calling the task compound splitting, and also in terms of identifying them from other words, calling the task compound identification. We begin by demonstrating this on Czech using deep learning and string matching. Then, on the same language, we generalize compound splitting task into parent retrieval, by building a tool called *Word Formation Analyzer for Czech*. It also covers derivation, meaning that we can trace an input word back to only a single word, and unmotivated words (recognizing that the input word has no ancestors) in addition to compounding. Finally, we present a multilingual parent retrieval and word formation classification tool called *PaReNT*, based around a custom-architecture deep model combining character-based and semantic representations, and show how the tool has been used in linguistic research.

We continue by applying this tool in combination with manual annotation to the Czech word-formation DeriNet, releasing version 2.2. We enrich this thus-far almost exclusively derivation-oriented data resource with information on compounding, and discuss the many considerations and decisions that were made along the way.

Finally, we survey the current coverage of compounds in Universal Dependencies in five languages (English, Czech, German, Dutch, Russian, Latin), and propose a way of modeling compounds by endowing them with a dependency structure and embedding them into the syntactic structure found therein.