LISTING AND DESCRIPTION OF COMPOUNDS —
METHODS AND APPLICATIONS.

PhD Dissertation
Agata Savary

This dissertation describes a natural language processing research in the field of nominal compounds in general and technical English. The starting point for the studies presented was INTEX, a tool for automatic treatment of large corpora.

While analyzing the problem of large coverage listing and describing of compounds, we addressed the following issues:

1) Which methods of compound description should be used?
2) For what kind of applications is this description useful?

The first issue is treated in the context of electronic lexical databases such as they are admitted in the INTEX system. We analyze the inflectional morphology of compounds in French, English and Polish. We propose a method of automatic generation of their inflected forms. We describe the construction of two electronic dictionaries: one for general English compounds, and the other for simple and compound terms of the computer science technical English. We also present a library of finite-state automata and transducers for the recognition of English cardinal and ordinal numerals.

The utility of large coverage compound dictionaries is verified through their application to two kinds of natural language processing tasks. First, we describe a method of acquisition of terms based on initial terminological resources. Secondly, we propose an automatic spelling checking algorithm of simple and compound words in a finite-state automaton dictionary.