Summary

The topic of this thesis is the phenomenon of factivity in contemporary Polish. The following three issues are primarily considered:

1. Which verbs in Polish are factive and which are not?
2. How often and in what way do Polish speakers use factive verbs?
3. How do computer models predict whether there is a relation of entailment, contradiction or neutrality between the whole utterance and its complement, and what linguistic data is required to improve their effectiveness?

To answer the first question, 325 primarily propositional verbs in contemporary Polish were analysed. Each verb was provided with an example of its use and information on whether it is factive.

The answer to the second question stems from the analysis of appropriately selected 2,348 utterances from the Polish Coreference Corpus (PKK), which is a representative sample of the National Corpus of Polish (NKJP). All utterances with the "V{że}p" structure found in the PKK were taken into account. An example of an utterance with such a structure is (1):

(1)   *Teraz już wiem, że był to zapach cebuli.* [NKJP]
        (*Now I know it was the smell of an onion.*)

What particularly interested me in such utterances was: (a) the type of main predicate around which the whole sentence is formed (factive / nonfactive) and (b) the relation between the whole sentence and the sentence complement that is part of it (entailment, contradiction and neutral).

To answer the last, third question, it was necessary to build appropriate computer models. Several such models were created as a result of a project conducted at the Faculty of Mathematics and Information Sciences of the Warsaw University of Technology. Their implementation, i.e. training and testing was based on the linguistic material prepared by me. Then, using the appropriate tools, I analysed their way of working, focusing primarily on determining what is necessary to improve their effectiveness.

The dissertation consists of five chapters. In the first chapter I focus on presenting an integrated conceptual and terminological grid that was used in the thesis. This grid is intended to be easily translatable into alternative conceptual constructs, and its possible translation should be irrelevant to the research results obtained.

I begin the second chapter by presenting the criteria I use to verify opposing hypotheses about the factivity of given lexical units. I then investigate whether a given unit is factive, moving on to a discussion of selected expressions. I have divided the analysed verbs into the following four semantic classes: *epistemic*, *perceptual*, *speech*, and *emotive* verbs. I devote by far the most attention to the epistemic verb *wiedzieć, że_* (*know that_*), which is the crucial unit in the class of factive expressions.

The third chapter primarily consists of the quantitative results of the analyses of PKK utterances. Among other things, the chapter presents data, such as the proportion of factive to nonfactive verbs, or the cooccurrence of nonfactivity and entailment.

In the fourth chapter, I focus on the results of machine learning models and I analyse their performance. I look at the models that were trained both on the basis of features in the dataset I created (e.g., the semantic class of the verb, its grammatical tense, the presence of internal negation) and solely on the basis of the whole utterance (so-called *word embedding*). The latter are based on the HerBERT model, a Polish version of the BERT model. The main task of the trained models was to predict one of three relations: entailment, contradiction or neutrality between the whole utterance and its complement *p*. I compare the output of the models, taking into account the results obtained for each of the three logical relations stated above.

The fifth chapter is a summary of the entire thesis. I return to the research questions posed and, in a summarized form, formulate and describe the obtained answers.