Different approaches
to the PP-attachment problem in Polish

Katarzyna Krasnowska
Institute of Computer Science
Polish Academy of Sciences
E-mail: k.krasnowska@phd.ipipan.waw.pl

Abstract
A number of approaches, using different available resources, were applied
to the PP-attachment problem in Polish. Some methods were reimplimenta-
tions of supervised and partially supervised models for English described
in literature, others were our modifications and extensions, mostly using
a wordnet for Polish. The best accuracy achieved on the final testing dataset
was 75.7%, which is not much below the accuracy of an expert’s decisions
obtained in a pilot study.

1 Introduction
The PP-attachment problem consists in identifying correct attachment sites for
prepositional phrases occurring in natural language utterances. A high-accuracy
method for solving this problem can be useful in parsing and parse disambiguation
for the purposes of creating treebanks as well as in any NLP application which
requires full syntactic analysis of text. The typical formulation of the problem’s
single instance is a quadruple \((v, n, p, n2)\), with verb \(v\) and noun \(n\) being two possible
attachment sites for a phrase headed by preposition \(p\)\(^1\) with a dependent noun
\(n2\). This work describes experiments on applying different approaches, using dif-
ferent available resources, to the PP-attachment problem in Polish.

2 Related work
A considerable amount of work has been devoted to the problem of PP-attachment,
especially in English. Extensive research in what could be called a “partially super-
vised” framework was started by Hindle and Rooth [10] and followed by, among

\(^1\) Polish has some prepositions which have the same surface form, but select for different gram-
matical cases and have different meanings. Therefore, throughout this text, unless explicitly stated
otherwise, by preposition we will mean its surface form together with the case.