In my talk, I will discuss the application of logistic regression techniques to the selection of linguistically relevant features for the determination of regular constructions that defy analyses in more traditional terms. The area of application will be Preposition Noun Combinations (PNCs) in German, as illustrated below.

(1) Die Aktion dient der Navigationsschulung über unbekanntem Gebiet.

The operation serves the navigational training above unknown territory.

PNCs ‘violate’ syntactic rules that require singular count nouns to appear together with a determiner (presumably to mark the singular). They are not entirely regular, but crucially regularly irregular, which is also reflected in their occurrence in a wide variety of languages that impose the same set of constraints on the realization of singular count nouns.

We will focus our study on two case studies concerning the prepositions ohne and unter. In the first case, PNCs seem to be more regular than PPs, while the second case exemplifies a closer relation to the expected patterns of regularity. The case studies show that features pertaining to the omission of determiners are partially lexically conditioned, but presumably also structural. The last part of the talk will address the role of logistic regression itself for feature selection. Several areas can be identified that cause potential problems or at least challenges for an analysis employing logistic regression: one is the underlying assumption that estimates are determined on the basis of maximum likelihood, another is the role of lexical influence in such models. With respect to the latter, we will compare the analysis to a generalized mixed linear model that indicates that at least in the case of some prepositions, PNCs are more likely to be triggered by lexical influence than by linguistic features.