

# An Algorithm for WSD Based on Collocational Data

---

Pascual Cantos,<sup>1</sup> Carlos Balsalobre, Vicenta Rodríguez  
and Ana María Fernández

## Abstract

This work presents a statistical model for meaning disambiguation based on a multivariate technique: discriminant function analysis. Succinctly, the discriminant function analysis is a multivariate statistical procedure of special interest in research as it allows to predict the probabilistic property of an individual to (a) group(s), previously established by means of discrimination functions.

The algorithm, we propose here, is the result of a previous, supervised discriminant function analysis performed for a polysemous word with  $n$  entries: as many entries as sample sentences/concordances found for all the meanings of that polysemous word. To each one of these  $n$  entries, we typified  $p$  independent numeric variables (*collocates*), which are like a feature profile of each one. Additionally, we added a further dependent quantitative variable (a classifying variable) with as many categories as meanings the polysemous word has. The result is a matrix sized  $n*(p+1)$ , where each case (sentence/concordance) appears with its profile and its allocation to a group (meaning). This final matrix is subject to contain all the knowledge used for later meaning discrimination of a polysemous word.

The poster shows the data of a meaning discrimination example, as well as a summary of all the algorithmic steps.

---

<sup>1</sup> *e-mail*: pcantos@um.es