

# A Program for Identifying Metaphor Candidates in Corpora

---

Tony Berber Sardinha<sup>1</sup>

## **Abstract**

The resources available for corpus-based metaphor analysts to go through a corpus looking for metaphor candidates are restricted to those generic tools available for corpus linguist research, such as concordancers, wordlisters, and frequency markedness identifiers. However, we know as yet very little about several essential properties of metaphors, such as what they look like, what their frequency is, what words seem to signal them, and so on, for a researcher to be able to enter a 'metaphor word' with confidence into a concordancer or tell which words are used metaphorically simply by looking at a frequency wordlist or keyword (words with marked frequency) list. It would seem, then, that metaphor identification in corpora is a task that computers might be well suited for, as they neither suffer from fatigue nor do they shift identification criteria along the way. A computer program that could screen an entire corpus and yield a list of potential metaphors would thus seem desirable. This would free the analyst from the tedious job of reading a whole corpus and marking up metaphors, or from running huge numbers of concordances only to find out that a minority of them actually returns metaphors. In this paper, I present an online metaphor identification program that identifies potential metaphors in Portuguese corpora. It does so by resorting to five databases that contain several kinds of information about lexis and its relationship to metaphor, all extracted from extensively hand-annotated corpora. These databases store the probability of words being metaphorical based on previous metaphor use, on lexical patterns occurring near and around words, and on their word class. In a preliminary evaluation, the program showed itself to be a valuable addition to the current metaphor analysts' corpus toolkit.

---

<sup>1</sup> *e-mail*: tony4@uol.com.br