A Proposal for Building the Knowledge Base of Onomasiological Dictionaries

Gerardo Sierra¹ and Laura Hernández²

Instituto de Ingeniería, Universidad Nacional Autónoma de México
¹gsierram@iingen.unam.mx
²lhernandezd@iingen.unam.mx

In this paper we present a methodology for creating and populating a lexical knowledge base (LKB) to be used in an onomasiological dictionary. The purpose of this methodology is to automatize the creation of specialized onomasiological dictionaries, which help to solve the “tip of the tongue” problem and assist authors in the active linguistic state (encoding). This article includes information about the main architecture of the LKB as well as our proposed dictionary. The methodology presented in this article allows the LKB to be populated with a wide variety of definitions from both colloquial and normative sources in such a way that by employing this LKB these specialized onomasiological dictionaries are able to handle users’ queries in natural languages.

Key words: onomasiology, dictionary, lexical knowledge base, computational linguistics, natural language processing

1. Introduction

People consult dictionaries for various purposes. Receptive linguistic users, readers or listeners, need a reference book to find the meaning of a word they have read or heard. However, in the active linguistic state, writers or speakers require wordbooks to satisfy the need to go from meaning or concept to a corresponding word.

Onomasiological dictionaries have been created precisely to help authors in a “tip of the tongue” state (Brown & McNeill, 1966), that is, a state in which they do not remember a word, or for authors who ignore the word expressing the meaning they have in mind.

©2011 Institute for Cognitive Science, Seoul National University
According to the kind of information contained, the structure, and the search strategy, onomasiological dictionaries can be classified as Thesauri, Synonym Dictionaries, Reverse Dictionaries, and Visual Dictionaries. Despite this great variety of options, people still have problems finding the correct word because these resources still assume too much from the user. For example, these dictionaries expect that users know the precise words to describe the object for which they would like to find the corresponding word. This is why some researchers have proposed to use natural language searching, also known as free-text searching (Lancaster, 1972). Onomasiological dictionaries allow users to input their descriptions of a concept or idea through more than one word in a search instead of using isolated words.

The main problem with this solution is that people express the same concept (say barometer) with different words (instrument, device, apparatus; determine, measure, record; etc.). As is shown below in Section 2.2, everyone thinks differently and there can be as many conceptualizations of the meaning of a word as there are people creating them. What is worse, these conceptualizations usually do not match the formal definitions found in conventional dictionaries.

One way to solve this problem is to obtain many descriptions from people with various backgrounds (academics, journalists, non-professionals, etc.). This would allow one to create a rich Knowledge Base likely to fit most users.

Since the Internet is the richest source of information available, it makes sense to use it to extract automatically term descriptions. To this end we use DESCRIBE, a tool to look for definitions of terms available on the web. DESCRIBE uses a set of Spanish language patterns commonly used when giving a definition. These patterns, known as Definitional Patterns, are classified in typographic and syntactic patterns. The first ones refer to the punctuation marks that usually connect the term with its definition, while the second ones allude to the verbs that introduce a defining predicate, such as define, know, denominate, etc.