

# Crowdsourcing Word-Color Associations

Mathieu Lafourcade<sup>1</sup>, Nathalie Le Brun<sup>2</sup>, and Virginie Zampa<sup>3</sup>

<sup>1</sup> LIRMM, Montpellier, France

<sup>2</sup> Imagin@t, Lunel, France

<sup>3</sup> LIDILEM, Grenoble, France

**Abstract.** In Natural Language Processing and semantic analysis in particular, color information may be important for processing textual information. Knowing what colors are generally associated with terms by people is valuable. We explore how crowdsourcing through a game with a purpose (GWAP) can be an adequate strategy to collect such lexico-semantic data.

**Keywords:** Word Color Associations, Lexical Network, Crowdsourcing.

## 1 Introduction

Color information is important in our daily life and may be of interest in the context of Natural Language Processing. Although this is beyond the scope of this paper, they are strong connections between colors and emotions. However, provide information about word-colors associations to a system dedicated to the semantic analysis of texts, in addition to other traditional knowledge (hypernym, parts of, semantic role, etc.) could greatly improve system performance. Association between word and color could be made for abstract nouns related to emotions (like fear, anger, danger, hope, ...) but in a more straightforward way for concrete nouns (like sky, lion, snow, sea, ...)

There is a very lively debate as to whether the associations between color and meaning were independent of age, gender, or nationality. This might be the case, for example, for the relation between *red* and *danger*, since the red may be related to *blood/fire* regardless of others factors. Berlin and Kay (1969) argued that differences could be organized into a coherent hierarchy, and that there is a limited number of *universal* basic color terms used in various cultures. Berlin and Kay based their analysis on a comparison of color words in 20 languages from around the world, but their findings have been discussed a lot. In the same way, several expressions using colors have the same meaning in different languages especially when they are culturally close, which is hardly a surprise. For example, *dark thoughts* in English and *idées noires* are roughly equivalent, as well as *to see red* and *voir rouge*. Many studies, mostly in English, have been undertaken to determine relations between colors and words or colors and emotions, etc. Most of those studies of psycholinguistic are undertaken in a classical way, and their raw results are general modest in size and not freely available. Furthermore, as previously mentioned, it is extremely delicate (and probably unwise) to translate directly the result of such studies from one language to another. Finally, we can say that there is no definite consensus on the universality of word-color associations for abstract or feeling words.

Mohammad (2011b) conducted experiments using eleven colors and showed that more 30% of the terms have a strong color association (for a lexicon of everyday words, not including specialized domains). About 33% of thesaurus categories (like Roget) have color associations, and abstract terms are associated with colors almost as often as physical entities do, mostly by metaphor. Again, Nijdam (2010) compares different models on relation between color and emotion and proposes a correspondence between emotions and colors. He concludes that some models about color-meaning may have some overlap but they also show a great amount of vagueness, certainly because the color is dependant of personal/cultural situation. In summary, the acquisition of data on associations between words and colors is hampered by high variability, (even restricted to a given language). The more abstract is the term, the higher the variation is to be anticipated. Gözde et al. (2011) make a resource that contains information about the association of words and colors in English. They made a short selection of 200 words, a subset of words used in Grefenstette (2005), and compare annotations made with Amazon Mechanical Turk service (10 annotators – 11 colors) with three automatic methods: image analysis, language models on web data and similarity between words and colors (using LSA).

At the present time, as far as we know, Gözde et al. (2011) are the only ones who tried to make a resource about association of colors and words in English. Such resources are very rare in other languages, and especially it doesn't exist in French language. But as noted by Grefenstette (2005) what people generally know about things (concrete things, but even abstract concepts) includes their typical color as an important component. That is why, this information is not usually shown in dictionary definitions or lexical resources, although it would be useful for various computer applications, and in NLP in particular. We think that color information can be very helpful in the context of automated lexical disambiguation (Word Sense Disambiguation - WSD). For example, the French word *tissu* is polysemous and can mean either *fabric* or *living tissues*. If in a given text, *tissu* is associated to *bleu* (blue), the color information can help to choose the right meaning and at least eliminate the wrong ones. Another computer application for which color information would be helpful is one that allows you to find a momentarily forgotten word but "on the Tip of the Tongue" (Lafourcade 2012 and Joubert 2012). The color, either actual or symbolic, is very often crucial for finding proper terms.

The objective of this paper is to present how to easily and efficiently produce a (French) resource that lists associations between colors and words, without resorting to corpora. For this purpose, we implemented a GWAP (game with a purpose as dubbed by L. von Ahn. (2006)) named ColorIt through which the player is asked to tell the color he spontaneously associates with a given term (<http://www.jeuxdemots.org/colorit.php>). At first we show how information on color is a valuable asset in the context of NLP, and then we detail the features of the game as well as some quantitative and qualitative results.

## 2 ColorIt, a Game for Collecting Colors

The goal of ColorIt is to collect spontaneous associations between colors and words, colors assigned to concrete terms as well as those symbolically and subjectively associated to verbs or nouns denoting abstract entities.