

Abstracts

ABSTRACT

- Coh-Matrix is a web-based application currently in development that automatically evaluates text. It uses two central concepts from discourse processing: text-based cohesion and situation-model based coherence. Cohesion is the degree to which components of the text are linked. Coherence is the representation of the world that the text conveys. Our intention is for Coh-Matrix to eventually map the cohesion of a text to the background knowledge and reading skills of the reader. Coh-Matrix will then be able to give feedback to a writer about which aspects of the text are cohesive and which lack cohesion. This will enable the writer to determine which aspects of the text need to be improved. Applications of Coh-Matrix on document quality as well as other future directions for the development of Coh-Matrix are discussed.

Abstract

- Text Complexity Analysis is an useful task in Education. For example, it can help teachers select appropriate texts for their students according to their educational level. This task requires the analysis of several text features that people do mostly manually (e.g. syntactic complexity, words variety, etc.). In this paper, we present a tool useful for Complexity Analysis, called Coh-Metrix-Esp. This is the Spanish version of Coh-Metrix and is able to calculate 45 readability indices. We analyse how these indices behave in a corpus of “simple” and “complex” documents, and also use them as features in a complexity binary classifier for texts in Spanish. After some experiments with machine learning algorithms, we got 0.9 F-measure for a corpus that contains tales for kids and adults and 0.82 F-measure for a corpus with texts written for students of Spanish as a foreign language.
- Key words:Complexity Analysis, Readability Assessment Indices, Coh-Metrix, Natural Language Processing

Abstract

- Concreteness ratings are presented for 37,058 English words and 2,896 two-word expressions (such as “zebra crossing” and “zoom in”), obtained from over four thousand participants by means of a norming study using internet crowd sourcing (передача некоторых функций кругу лиц без подписания трудового договора) for data collection. Although the instructions stressed that the assessment of word concreteness would be based on experiences involving all senses and motor responses, a comparison with the existing concreteness norms indicates that participants, as before, largely focused on visual and haptic experiences. The reported dataset is a subset of a comprehensive list of English lemmas and contains all lemmas known by at least 85% of the raters. It can be used in future research as a reference list of generally known English lemmas.

Abstract

- We evaluated the effectiveness of new indices of text cohesion to determine the appropriate human assigned grade level of a text. In particular, we investigated the 'efficacy of automated text indices produced by the online tool Coh-Matrix in predicting the grade level assigned by publishers to their own textbooks. To do this, we sampled 311 school textbooks from a large database, choosing roughly equal numbers of science, narrative, and social science texts. Publisher-assigned grade levels were found to be moderately predictable by traditional approaches such as the Flesch-Kincaid Grade Level. Prediction of grade level was significantly improved by the inclusion of cohesion indices obtained by Coh-Matrix. Implications for the improvement of textbook selection are discussed.

- The vast majority of brain-injured patients with semantic impairment have better comprehension of concrete than abstract words. In contrast, several patients with semantic dementia (SD), who show circumscribed (ограниченная) atrophy of the anterior temporal lobes (передние височные доли) bilaterally, have been reported to show reverse imageability effects, that is, relative preservation of abstract knowledge. Although these reports largely concern individual patients, some researchers have recently proposed that superior comprehension of abstract concepts is a characteristic feature of SD. This would imply that the anterior temporal lobes are particularly crucial for processing sensory aspects of semantic knowledge, which are associated with concrete not abstract concepts. However, functional neuroimaging studies of healthy participants do not unequivocally predict reverse imageability effects in SD because the temporal poles sometimes show greater activation for more abstract concepts. The authors examined a case-series of 11 SD patients on a synonym judgment test that orthogonally varied the frequency and imageability of the items. All patients had higher success rates for more imageable as well as more frequent words, suggesting that (1) the anterior temporal lobes underpin (поддерживать) semantic knowledge for both concrete and abstract concepts, (2) more imageable items—perhaps because of their richer multimodal representations—are typically more robust (крепкий) in the face of global semantic degradation and (3) reverse imageability effects are not a characteristic feature of SD.

Abstract

This paper investigates the variation in cohesion across written and spoken registers. The same method and corpora were used as in Biber's (1988) study on linguistic variation across speech and writing; however instead of focusing on 67 linguistic features that primarily operate at the word level, we compared 236 language and cohesion features at the text-level. Variations in frequencies across these features provided evidence for six dimensions: (1) speech versus writing, (2) informational versus declarative, (3) factual versus situational, (4) topic consistency versus topic variation, (5) elaborative versus constrained, (6) narrative versus non-narrative. Our cohesion and linguistic analysis showed most variation in speech and writing, whereas the linguistic feature analysis operating at the word level did not yield any difference.

Introduction

Abstract

- In this paper we present the PorSimples project, whose aim is to develop text adaptations tools for Brazilian Portuguese. The tools developed cater for both people at poor literacy levels and authors that want to produce texts for this audience. Here we describe the tools and resources developed over two years of this project and point directions for future work and collaboration. Since Portuguese and Spanish have many aspects in common, we believe our main point for collaboration lies in transferring our knowledge and experience to researches willing to develop simplification and elaboration tools for Spanish.