

# Web Interface to Measure Language Abstraction Using the Linguistic Category Model

By Andrew Gallant

## Introduction

The Linguistic Category Model (LCM) is a model about the psychological properties of interpersonal language, which is used to classify verbs and adjectives that represent actions and states between people. The goal of this project was to develop an online interface such that the user can apply the LCM to any number of documents. The online interface makes use of keyboard shortcuts, AJAX, and a relational database to help speed up the process of applying the LCM, and consequently retrieve results more efficiently.

The LCM works by assigning each adjective and verb in a target text a designated score. Verbs are split up into four distinct categories: descriptive action verbs, interpretative action verbs, state action verbs, and state verbs. With adjectives being the most abstract and descriptive action verbs being the most concrete, the scores are then used to calculate the final LCM score which measures how abstract (high score) or how concrete (low score) the target text is.

## Problem

For this particular project, the researcher that this application was built for was examining annual shareholder reports, measuring the LCM score of each report, and comparing it to other market related data (i.e., how well the company is doing). In doing so, each report had to be hand coded on paper and LCM scores manually tabulated. Given that such a research project requires analyzing hundreds of documents, a faster approach was greatly desired.

## Solution

After doing some background research on available linguistic related programs, it was clear that none of them suited the very specific requirements of the LCM. Thus, the logical proceeding was to move from hand coding documents

to coding them on a computer. The human process of coding documents could not be cut out, but the procedure could be dramatically quickened through technological aide, and tabulating the LCM score of each document could be done automatically. (Additionally, different dimensions of data collected could be viewed with a single click.)

Since any research involving human decision-making needs protection against bias, each document had to have the ability to be coded by more than one person- with each person only able to see their coding. From each person's coding, an inter-rater reliability score using Cohen's Kappa coefficient can be generated.

Given the aforementioned conditions, a web interface was the logical choice. Only one copy of each document would have to be stored, and calculations depending upon more than one person's coding could be automatically tabulated, making data comparison decidedly easier.

## An Example

Jeannette Schmid of Universität Heidelberg researched media coverage of show wrestling. More specifically, show wrestling consists of a separation between morally bad fighters who win by cheating and morally good fighters who win because of their intrinsic goodness. Using the LCM, Schmid analyzed published fight reports in an official wrestling magazine to show that badness and goodness can be attributed on a stable and global level and that fights with a predictable outcome are described in language related to the righter in contrast to those fights with an unpredictable outcome.

## The Application



Home screen

Once a document is added to the system, it is ready to be scored. The relational database design is fairly straightforward with one caveat: how to store each individual document. Given that each and every word must be able to have its own specific coding, the logical design is to create a table that keeps a record for each individual word of each document. Thus, once a document is added, its source text cannot be modified.

## Coding

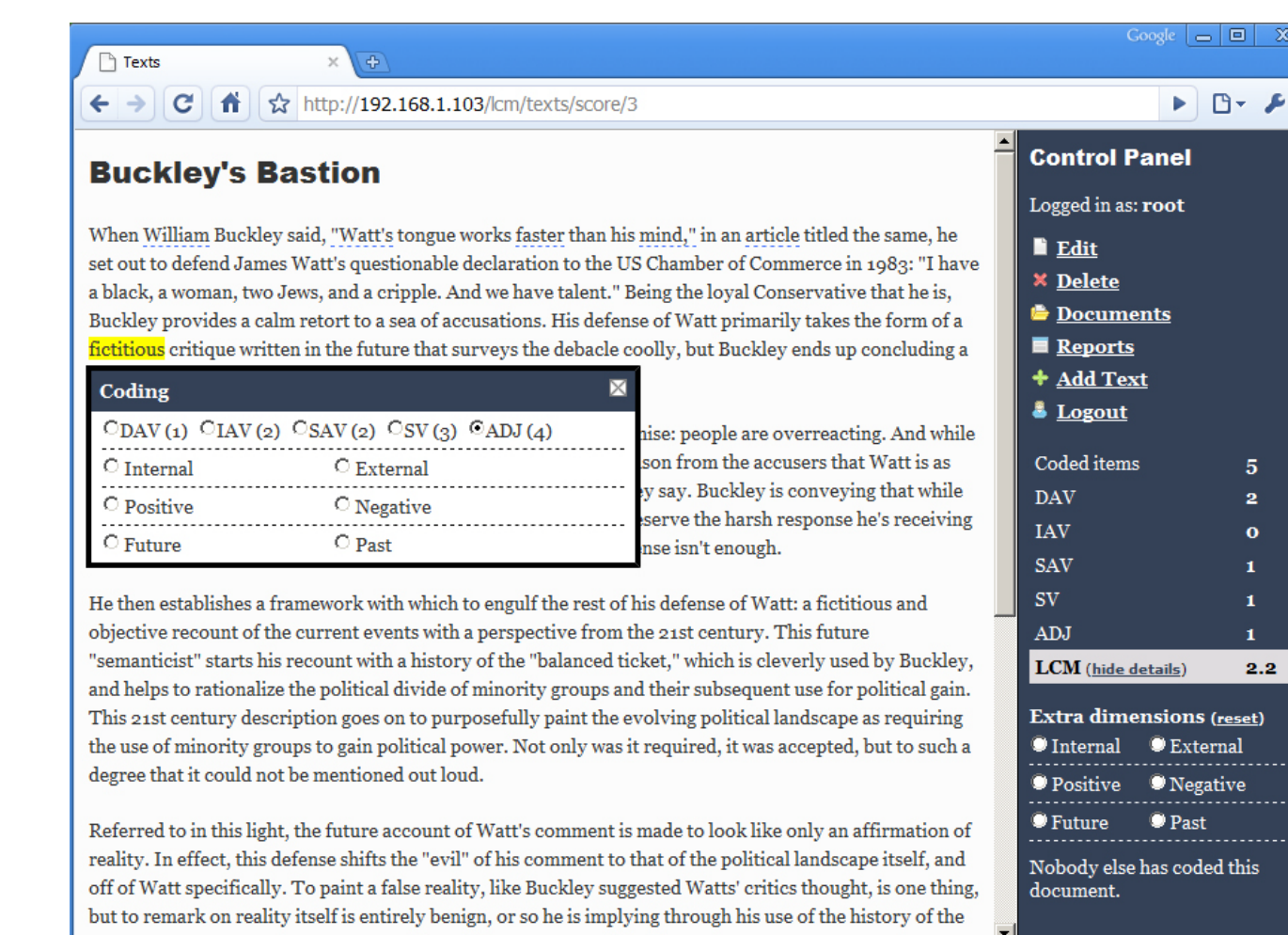
The coding screen for each document presents the user with the source text, along with a control panel of options on the right side used for navigation and keeping track of the LCM score in real time. The control panel allows the user to view the individual breakdown of each scoring category (adjectives and four different types of verbs), along with any additional dimensions that were specified when the document was added.

The coding screen of the document allows the user to highlight one or more words, from which a coding dialog appears. (One or more words is allowed, as sometimes a phrase needs to be classified as one type of verb or adjective.) This method of coding is usable with a

mouse, but can also be quickly accomplished using familiar shortcut keys (such as holding shift to highlight multiple words). Codings can be modified or deleted, too. With the data generated by user codings, LCM scores and other data can be downloaded in excel format.

## Future Possibilities

The next logical step from this application is to remove the manual human element of coding documents. While this is currently being investigated, it requires further advanced research in computational linguistics as it relates to the LCM.



Coding screen