

Chapter 95

The Gramulator: A Tool to Identify Differential Linguistic Features of Correlative Text Types

Philip M. McCarthy

The University of Memphis, USA

Shinobu Watanabe

The University of Memphis, USA

Travis A. Lamkin

The University of Memphis, USA

ABSTRACT

Natural language processing tools, such as Coh-Metrix and LIWC, have been tremendously successful in offering insight into quantifiable differences between text types. Such quantitative assessments have certainly been highly informative in terms of evaluating theoretical linguistic and psychological categories that distinguish text types (e.g., referential overlap, lexical diversity, positive emotion words, and so forth). Although these identifications are extremely important in revealing ability deficiencies, knowledge gaps, comprehension failures, and underlying psychological phenomena, such assessments can be difficult to interpret because they do not explicitly inform readers and researchers as to which specific linguistic features are driving the text type identification (i.e., the words and word clusters of the text). For example, a tool such as Coh-Metrix informs us that expository texts are more cohesive than narrative texts in terms of sentential referential overlap (McNamara, Louwerse, & Graesser, in press; McCarthy, 2010), but it does not tell us which words (or word clusters) are driving that cohesion. That is, we do not learn which actual words tend to be indicative of the text type differences. These actual words may tend to cluster around certain psychological, cultural, or generic differences, and, as a result, researchers and materials designers who might wish to create or modify text, so as to better meet the needs of readers, are left somewhat in the dark as to which specific language to use. What is needed is a textual analysis tool that offers qualitative output (in addition to quantitative output) that researchers and materials designers might use as a guide to the lexical characteristics of the texts under analysis. The Gramulator is such a tool.

DOI: 10.4018/978-1-4666-4979-8.ch095

WHAT IS THE GRAMULATOR?

The *Gramulator* is a freely available qualitative and quantitative computational textual analysis tool. It is designed to allow researchers and materials designers to identify indicative lexical features of texts and text types. More formally, the Gramulator is designed to *identify differential linguistic features of correlative text types*.

The Gramulator primarily functions by analyzing two *sister corpora*. Sister corpora are two highly related, yet theoretically distinguishable, sets of data. The Gramulator analyzes these two sister corpora relative to each other, and identifies the lexical features (called *differentials*) that are indicative of one sister corpus while being antithetical of the other sister corpus. Thus, the Gramulator always produces two sets of output (i.e., differentials), and these sets of differentials are always relative to the two sets of data analyzed.

As examples of suitable sister corpora for the Gramulator to process, *science fiction literature* could be compared to *romantic literature*, *modernist poetry* could be compared to *post-modern poetry*, *Shakespearean drama* could be compared to *Greek tragedy*, or *introduction sections* from scholarly articles could be compared to *discussion sections*. The object of such analyses is to *shake out* from the two sister corpora (i.e., the *correlative corpora*) the lexical features that differentiate, distinguish, and identify the respective text type (i.e., produce the differentials).

THE RELEVANCE OF THE GRAMULATOR

Identifying indicative features of text types is important to language learners, educators, material designers, and a wide range of researchers from fields such as cognitive science, linguistics, and second language learning. Explicit knowledge of these indicative lexical features helps in several ways. For example,

- A better understanding of language conventions and practices can be attained, especially when and where two closely related registers meet (e.g., reviewing and summarizing);
- Materials to meet the needs of students can be better prepared, especially where focused practice is called for (e.g., for language learners);
- A better understanding of register hierarchies to distinguish cases of different practices from cases of advanced practices can be achieved;
- The characteristics of the moves that constitute the functionality of registers can be identified; and
- Differences between similar texts (e.g., a manipulated text) can be accessed more easily so that potential learning gains can be better established.

Looking at these issues in more detail, we begin with language conventions. The use of indicative lexical features that constitute a genre, register, or variety (or broadly speaking a *text type*) are typically unconsciously agreed upon conventions (Downs 1998; Hymes 1972). Having knowledge of these conventions is beneficial to those who need or want to diversify their discourse communities because such knowledge provides a greater likelihood of the employment of appropriate comprehension strategies and production formats. At a finer level, awareness of discourse specific language features is facilitative for memory activations, expectations, inferences, depth of comprehension, evaluation of truth and relevance, pragmatic ground-rules, and other psychological mechanisms that depend upon discourse interpretation (Bhatia 1997; Graesser et al. 2002; VanDijk & Kintsch 1983; Zwaan 1993).

Turning to pragmatic issues, teachers' and materials designers' understanding of indicative features of text types offers the possibility of better preparing texts that are appropriate for the designated audience. For example, in language learning

20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/the-gramulator/97556

Related Content

Rewriting of Text and Paratext: Reception of "Bushido: The Soul of Japan" in a Chinese Context

Xiao Li (2022). *International Journal of Translation, Interpretation, and Applied Linguistics* (pp. 1-12).

www.irma-international.org/article/rewriting-of-text-and-paratext/304076

How Does Gamified L2 Learning Enhance Motivation and Engagement: A Literature Review

Mourad Majdoub and Géraldine Heilporn (2024). *Fostering Foreign Language Teaching and Learning Environments With Contemporary Technologies* (pp. 134-173).

www.irma-international.org/chapter/how-does-gamified-l2-learning-enhance-motivation-and-engagement/336058

On the Construction and Application of a Platform-Based Corpus in Tourism Translation Teaching

Beibei Lin and Po-ching Yip (2020). *International Journal of Translation, Interpretation, and Applied Linguistics* (pp. 30-41).

www.irma-international.org/article/on-the-construction-and-application-of-a-platform-based-corpus-in-tourism-translation-teaching/257028

Perspectives on English Language Learners: Addressing Effective Pedagogical Practices for Bilingual Early Childhood Students' Reading Success in Dual Language Classrooms

Isela Almaguer and Michael Whitacre (2018). *Handbook of Research on Pedagogies and Cultural Considerations for Young English Language Learners* (pp. 325-343).

www.irma-international.org/chapter/perspectives-on-english-language-learners/190986

Attention-Sharing Initiative of Multimodal Processing in Simultaneous Interpreting

Tianyun Li and Bicheng Fan (2020). *International Journal of Translation, Interpretation, and Applied Linguistics* (pp. 42-53).

www.irma-international.org/article/attention-sharing-initiative-of-multimodal-processing-in-simultaneous-interpreting/257029