Chapter 19
From Words to Integers and Beyond in Corporate Life

Boris Kabanoff
Queensland University of Technology, Australia

Mahani Hamdan
Universiti Brunei Darussalam, Brunei

ABSTRACT
Drawing upon the chapters from this volume for exemplars, this chapter provides a critical appreciation of Computer-Aided Text Analysis (CATA), a method that analyses texts that occur naturally or are evoked by the researcher, where the underlying approach is usually positivist in nature and the focus is on studying the presence of certain concepts that are inductively or deductively derived, with software playing a key role in identifying and quantifying the presence of concepts in text. The first section considers some of the major characteristics and differences in CATA approaches to issues such as latent vs. manifest content and the interpretation of the content of communications from either a tone or attention perspective. The second considers enduring, widely accepted assumptions about CATA that centre upon the existence of intimate associations between words and underlying psychological constructs, and between word frequencies and the importance or centrality of constructs in cognitive processes. The final section considers enduring challenges and opportunities including the complex and vexed issues of impression management and textual context. The authors conclude that in relation to the study of the language of corporate life, CATA can be viewed as a glass half-full rather than half-empty.

INTRODUCTION
Language is both a lock and a key (Chinese poet Yan Zhen quoted in Uren, “The People’s Poet”, 2013, p. 28.)

Poets have a licence to couch great truths in succinct, emotionally powerful, and perhaps slightly mysterious and ambiguous ways. On the other hand, it is the task of academics to explore such truths intellectually, in depth and detail, identifying the key constructs and their underlying relations and structures, hopefully without impairing the essential truth. So it could be said that in Janu-
ary 2013, around 60 academics gathered at the University of Texas, Austin under the benign and encouraging eye of their own muse, Professor Rod Hart, to play their role in exploring and explaining the underlying truth of Yan Zhen’s words.

The goals of this chapter are quite broad. Rod was explicit and yet also somewhat Delphic in his expectations and aspirations for the chapter. Even though DICTION was a key analytic tool in most chapters, this chapter was not to be about DICTION per se, or simply a critique of the individual chapters forming this section of the book. Rather DICTION and these studies, as well as some others that got our attention, were to be more a launching pad for observations on what they revealed about the current state of understanding and research into the language of institutions, as well as some ‘adventurous’, but not too outlandish reflections on future challenges and opportunities.

So the bulk of our chapter is organised under three, roughly descriptive headings: The Launchpad: An Initial Appreciation, Enduring Assumptions, and Enduring Challenges and New Frontiers. The next section therefore offers an overall appreciation of some of the key characteristics of these chapters as a whole, that is, it is a bit of a reconnaissance of the launch pad, and provides some guidance to the launch trajectory for subsequent sections.

**THE LAUNCHPAD: AN INITIAL APPRECIATION**

**Computer Aided Text Analysis (CATA)**

We think it is reasonable to describe virtually all the chapters here as examples of CATA. What defines CATA? Without viewing the following as a comprehensive definition, generally, CATA analyses texts that occur naturally (e.g. annual reports, speeches, media headlines or stories) or are evoked by the researcher (e.g. interviews that are converted to text). The underlying approach is usually *positivist* in nature and the focus is on studying the presence of certain concepts that are inductively or deductively derived; and the software plays an important role in identifying and quantifying the presence of concepts in text though the role of quantification ranges from being the key for interpreting text to assisting qualitative interpretation (c.f. Kabanoff, 1997; Pollach, 2012).

In our own work, quantification has always played a central role as it does in the vast majority of chapters here. We think there is something to be said for restricting the term CATA to research in which quantification is a key objective and basis for interpretation, while using an alternative term such as computer–aided interpretive text analysis (CAITA) for work where quantification plays a secondary, supporting role. This would be useful for signalling the different ‘text analysis tribes’ we belong to. In the first author’s case for a start, it would reduce the frequency with which research students come to him, to his amusement and their frustration, announcing they are there because they have heard “I am an expert in qualitative analysis”. However, it is worth remembering that even when one uses the most software-driven form of CATA involving applying pre-existing, general dictionaries for scoring text, there is an important role for the researcher to provide evidence for the validity of the scoring procedure by ‘returning to the text’ (Weber, 1990). This can involve formal procedures for assessing validity and reliability using one or more human coders, or at a minimum the researcher reading examples of text that have been scored by the software to try to ensure that the scoring process is doing what it claims to be doing. Perhaps there will come a time when we can assume some software’s validity and reliability but we do not think we are there yet. So it was pleasing to see in these chapters many examples of researchers ‘returning to the text’ to support the automated text scoring process.

Instead of CATA, we could have used alternative terms such as computer aided content analysis