CHAPTER 11

QUANTITATIVE TEXT ANALYSIS

Matthias R. Mehl

Text analysis has been receiving increasing attention within the social sciences. This surge of interest is reflected in several recent books (Neuendorf, 2002; Popping, 2000; Smith, 1992; Weber, 1990; West, 2001), chapters (e.g., Smith, 2000), and review articles (e.g., Lee & Peterson, 1997; Pennebaker, Mehl, & Niederhoffer, 2003) on the topic. This chapter seeks to demonstrate that quantitative text analysis is a powerful, efficient, and easy-to-use scientific method with a wide spectrum of applications in psychology. It is organized into four major sections. At the beginning, an example of text analysis from social psychology is presented. This is followed by a brief historical overview of text analysis and an introduction to the conceptual foundation of text analysis. The third section reviews nine influential text analysis approaches in psychology. The final part discusses potentials and problems of quantitative text analysis.

A TEXT ANALYSIS EXAMPLE FROM SOCIAL PSYCHOLOGY

How do people respond to physical symptoms, and what makes them decide whether or not to seek treatment? The methodological toolbox in psychology is large, and there are a number of potential ways to address this question. Yet the default strategy has been to rely on just one tool—the questionnaire. In this case, for example, a researcher might create a health-decision questionnaire consisting of a number of Likert-scaled items, such as "How serious do your symptoms have to be before you see a doctor?" and "When you experience symptoms, how long do you wait before you see a doctor?" An alternative approach is simply to ask people what they normally do when they experience some rather common physical symptoms.

In a recent introductory psychology class, students wrote brief essays on how they would react if they woke up sweating, feeling terrible with a 102°F fever, and having a rash on their chest. Consider the following three responses[1]:

> *Participant A:* My initial impressions would be panic, going through heightened anxiety. Health is probably my highest priority here at the university, and any slight deviation from feeling decent would send off warning signals to get help ASAP. Initially, I would go to my primary source of 24/7 counseling: calling home. They wouldn't mind at all. Calling them would give me a good idea of what I might be coming

Preparation of this chapter was aided by a grant from the National Institutes of Health (MH52391) to James W. Pennebaker. I am grateful to Sherlock Campbell, Michael Eid, Samuel Gosling, James Pennebaker, Lisa Trierweiler, and an anonymous reviewer for their comments on previous drafts of this chapter.

[1] I thank Carla Groom and James Pennebaker for providing the essays.

> down with. I have my own physician's number at hand, and if the symptoms persisted throughout the rest of the morning, I wouldn't be hesitant as to calling him.
>
> *Participant B:* I would first call my mother and tell her about my situation. I would see what she would suggest, which would most likely be to go see a doctor. I would call the University Health Center and make an appointment to see a doctor that day. Because I am covered by my mother's health insurance, the co-pay for me visiting the doctor would be twenty dollars. If the doctor knows what is wrong with me and gives me a prescription, the twenty dollars would be well spent.
>
> *Participant C:* First thing I would do is try and remember if I had ever experienced similar symptoms so I could try to figure out on my own what was wrong with me. I would then probably call my mother to see if she had any idea what could be causing my symptoms and if she thought I should see a doctor. Knowing me, I would worry myself into a panic attack if I let the symptoms persist since I do not like not knowing what is wrong with me. I have gotten sick so often during the past few years that I have given up on trying to just cope with any sort of illness by myself.

What is striking about these answers is that, on the surface, all three participants reacted quite similarly. They all say they would go to see a doctor on the first day. For all three participants, one of the first things they thought about was calling their family. They probably also didn't differ much in terms of how serious they considered their symptoms to be. Thus, their responses to a multiple-choice questionnaire would most likely be comparable. However, a quick read of their responses conveys impressions of psychological reactions that are quite distinct.

For example, Participant B adopted a rather cool and rational attitude, compared to Participants A and C, who reacted rather emotionally. The free responses also tell us that health is clearly an important—almost dramatic—factor in Participant A's life, whereas economic considerations prevail in Participant B's thinking. Finally, there is a sense that Participant C is somewhat self-preoccupied and slightly socially isolated. It is likely that these differences—although not having an immediate impact on whether or not to see a doctor—ultimately translate into behavior relevant to the researcher's question (e.g., in terms of their expectations of the doctor or compliance with a prescribed treatment).

Of course, ad hoc impressions always run the risk of being subjective. A text analysis program such as *Linguistic Inquiry and Word Count* (LIWC; Pennebaker, Francis, & Booth, 2001) can paint a more objective picture. LIWC calculates the percentage of words that falls into a number of grammatical (e.g., pronouns, articles, prepositions) and psychological (e.g., words indicating emotional, cognitive, or social processes) categories. As shown in Table 11.1, LIWC analyses of the three essays generally support our intuitions: Participant C indeed used fewer emotion words than Participants A and B, and the considerably lower rate of social words and the frequent use of first-person-singular self-references (I, me, my) support our hunch that Participant C is less socially integrated and more self-absorbed than the other two students.

The LIWC analyses, however, reveal more than meets the eye: Participant A has a tendency to use long words (a marker of cognitive complexity); Participant B uses articles at a high rate (a marker of a concrete thinking); Participant C's writings contained a large number of cognitive words (a marker of mental processing). The three also differ in other important ways, such as their orientations to time (Participant A, B, and C: future, present, and past tense, respectively). Thus, a simple word count analysis provides insights into the participants' psychological worlds that go far beyond what multiple-choice questionnaires typically capture.

TABLE 11.1

Linguistic Inquiry and Word Count (LIWC) Analysis of Three Participants' Answers to the Question, "How Would You React If You Woke Up with a Series of Physical Symptoms?"

LIWC Variable	Participant A	Participant B	Participant C
Total word count	100.0	89.0	116.0
Words of more than six letters	25.0	12.4	10.3
First-person-singular pronouns	9.0	11.2	15.5
Articles	5.0	10.1	3.5
Prepositions	15.0	7.9	16.4
Emotion words	5.0	1.1	3.5
Positive emotion words	2.0	0.0	0.0
Negative emotion words	3.0	1.1	3.5
Cognitive mechanisms	10.0	16.9	21.2
Social processes	9.0	10.1	4.3
Past-tense words	0.0	1.1	5.2
Present-tense words	6.0	14.6	10.3
Future-tense words	4.0	3.4	0.9
School-related words	2.0	1.1	0.0
Money-related words	0.0	4.5	0.0

Note. All LIWC variables except total word count are expressed in percentages of total words.

TEXT ANALYSIS AS A SCIENTIFIC METHOD: HISTORICAL OVERVIEW AND CONCEPTUAL FOUNDATION

As a scientific method, text analysis is still young. It experienced its first surge during World War II, when Allied governments launched a series of large-scale projects to analyze the content of Nazi propaganda (Krippendorff, 1980). Stimulated by Murray's (1938) work on the Thematic Apperception Test (TAT), the first postwar decade in psychology was characterized by an avalanche of studies on the assessment of implicit motives via thematic content analysis (Smith, 1992). The advent of mainframe computers in the early 1960s revolutionized the field. Stone and his colleagues at Harvard University developed the first computerized text analysis program: the General Inquirer (Stone, Dunphy, Smith, & Ogilvie, 1966).

Since the 1970s, scientific text analysis has been shaped by two other technological advancements: the diffusion of personal computers with exponentially growing processor speeds and the rapidly increasing digitalization of data—through the Internet and progress in optical character and voice recognition (West, 2001). Computers have become increasingly sophisticated word search engines and, most recently, have been used for extracting semantic and grammatical relationships among words (Foltz, Kintsch, & Landauer, 1998; Roberts, 1997).

Defining Text Analysis

Not surprisingly, there has been disagreement on how to define text analysis. Shapiro and Markoff (1997) suggest the following minimal definition: Text analysis is "any systematic reduction of a flow of text (or other symbols) to a standard set of statistically manipulable symbols representing the presence, the intensity, or the frequency of some characteristics relevant to social science" (p. 14). This definition includes both qualitative (Riessman, 1993; Shiffrin, 1994) and quantitative approaches. In accord with the notion of measurement in this handbook, this chapter focuses exclusively on quantitative text analysis applications.

Classification of Quantitative Text Analysis Approaches

Quantitative text analysis approaches vary along a variety of different dimensions (Popping, 2000; Robins, 1997; Smith, 1992). The following section introduces four conceptual distinctions that provide a framework for organizing the existing approaches in psychology.

Aim: representational versus instrumental. On the broadest level, text analysis methods differ with regard to whether they are representational or instrumental in aim (Popping, 2000; Roberts, 1997). The role of the receiver in normal communication is to decode as accurately as possible the intended meaning of a message. This is what representational text analysis seeks to achieve. Its goal is to develop a representation of the sender's original intention of a message. In doing so, representational analysis is interested in the manifest content of a text.

Instrumental analyses focus mainly on latent content. Independent of the author's intention, a message is analyzed for occurrences of a set of themes (e.g., hostility, anxiety, need for power). The linguistic analysis at the beginning of the chapter, for instance, was instrumental because—rather than representing what the students intended to say—it focused on selected psychological aspects of language use (e.g., words hinting at emotional and social functioning).

So far, most existing text analysis applications in psychology have been instrumental. Compared to other sciences, psychology is highly deductive in its research. Instrumental analyses allow the specification of linguistic variables as the operationalizations of theoretical constructs and thus facilitate hypothesis testing. Also, psychology has a history of going beyond manifest content by reading between the lines to unravel the "unspoken" yet psychologically existing meaning—a task that only instrumental analyses accomplish. Finally, instrumental analyses can be performed on any desktop computer; a representational analysis' mimicking of natural syntax is computationally intensive and generally requires specialized machines (as well as users).

Approach: thematic versus semantic. The second conceptual distinction concerns the extent to which text analysis exclusively identifies themes or also models the relationships among them (Popping, 2000; Roberts, 1997). Until the 1980s, virtually all text analysis was thematic in nature. Thematic text analysis maps the occurrence of a set of concepts in a text and thus can technically be solved by counting the frequency of particular target words or phrases.

Semantic text analysis seeks to extract information on the conversational meaning of a theme. For example, it can be crucial to know not only that the theme "killing" is mentioned in a text but also whether it occurred in the context of "self" or "other people." Semantic text analysis solves this problem by specifying the concrete nature of relations among themes. Hence, the level of analysis in the semantic approach is typically the clause. Semantic text analysis first specifies a semantic grammar, a subject–verb–object (S-V-O) template, in which the concepts of interest are arranged like pull-down menus (e.g., [I/we] or [he/she/they] or [an object]; [S]–killed [V]–the dog [O]). It then determines the frequency with which certain concept constellations occur. In the example at the beginning of the chapter, semantic analysis could, for example, determine how often students call their mother and go to the doctor on her recommendation—as compared to the mother calling the student or the student calling the mother after returning from the doctor.

Recently, a new development in the field, latent semantic analysis (LSA), has received an increasing amount of attention (Folz et al., 1998; Landauer & Dumais, 1997). Compared to traditional semantic approaches where an investigator defines the context in a "top-down" manner, LSA constitutes a "bottom-up" approach, where information about the semantic similarity of words is extracted by analyzing their usage across a large body of texts. Because of its flexibility, computational power, and conceptual similarity to human cognition, it is a tool with great potential for the area of psychology (Campbell & Pennebaker, 2003).

In allowing the identification of themes *and* the relations that exist among them, semantic text analysis provides an additional degree of freedom. For evaluating its overall effectiveness, however, it is important to keep in mind that the meaning of a sentence is rarely revealed in its surface grammar. A powerful semantic analysis thus would need to identify the underlying deep structure—a task that is yet impossible to delegate entirely to a computer. Consequently, most semantic text analysis relies on human coders to parse large amounts of texts (Popping, 2000).

Bandwidth: broad versus specific. Text analysis approaches also differ in their bandwidth (Pennebaker et al., 2003). Some approaches focus on less than a handful of specific linguistic variables. Mergenthaler (1996), for example, analyzes therapy protocols exclusively for a client's use of emotion words and cognitive words and ignores other potentially relevant information, such as the content of the therapy session or a client's linguistic style. Other approaches intend to provide a broad linguistic profile of a text. LIWC, the text analysis program from our initial example, for instance, measures up to 82 grammatical and psychological language parameters.

Although specific approaches tend to have a stronger theoretical background, broad approaches usually are more inductive and phenomenon oriented. Researchers who find a text analysis program that captures exactly what they are interested in might prefer it to an "all-rounder" type of software because of its supposed better power. However, in those cases where a compromise needs to be made between what one is interested in and what is "out there," applications with broader bandwidth offer more flexibility.

Focus: content versus style. The fourth distinction concerns the "what" versus "how" in text analysis (Groom & Pennebaker, 2002). Conceptually, it dates back to Allport's (1961) distinction between adaptive and stylistic aspects of behavior. Whereas the adaptive components of a behavior are intended and purposeful in a given context (e.g., initiating a conversation), its stylistic aspects are mostly unintended, automatic, and serve expressive rather than instrumental functions (e.g., nervous gestures while initiating the conversation). Applied to verbal behavior, this distinction captures the difference between *why* a person is saying something, that is, the content of a statement (e.g., "When does the next number 5 bus pass by?"), and *how* the person is saying it (e.g., "Excuse me, would you possibly know when the next number 5 bus is supposed to pass by here, please?"). Looking "behind" a message for verbal mannerisms (Weintraub, 1981) or linguistic styles (Pennebaker & King, 1999) reveals more subtle aspects of a communication.

Historically, both strategies have been successful in psychology (Pennebaker et al., 2003; Smith, 1992). What makes stylistic language analyses particularly intriguing is that humans naturally attend to what people are saying or writing. It is cognitively quite demanding to tune out the meaning of a message for the sake of attending to particularities in word choice (cf. Hart, 2001). Consequently, for human judges linguistic styles are hard to detect and thus constitute the perfect target for computerized word count programs that are blind to meaning.

Summary. Conceptually, text analysis applications can be organized according to whether they are representational or instrumental in their aim, thematic or semantic in their approach, broad or specific in bandwidth, and focused on language content or style. Although these distinctions may not always be clear in practice, they offer a heuristic framework for deciding which text analysis strategy to use for a certain kind of research question. The following section provides a more concrete picture of how text analysis has been applied in psychology.

QUANTITATIVE TEXT ANALYSIS APPROACHES IN PSYCHOLOGICAL RESEARCH

This section reviews nine quantitative text analysis approaches that have been highly influential in psychology. The approaches were selected to be reasonably representative of the spectrum of existing text analysis strategies. More comprehensive reviews can be found in Popping (2000), Roberts (1997), Smith (1992), and Pennebaker et al. (2003). For each method, the historical and theoretical background is provided along with a description of how text samples are analyzed. Finally, each approach is located within the four-dimensional conceptual framework introduced in the previous section. Table 11.2 provides an overview of the depicted approaches. The approaches are presented roughly in order of historical development.

Thematic Content Analysis

Thematic content analysis is used here as a summary label for a number of approaches that have been developed in the context of motivational psychology (Smith, 1992). Generally, these approaches have human judges identify critical thematic references in a text. Ratings are made either each time a theme occurs or as global ratings reflecting the prevalence of a theme across an entire text. In either case, the analyses are based on standardized coding systems that define a psychological construct by specifying rules for when a certain theme is and is not considered indicative of the construct. Judges undergo extensive training until a predefined degree of agreement is obtained. Smith's (1992) *Motivation and Personality: Handbook of Thematic Content Analysis* contains detailed descriptions of 14 different coding systems. The following section highlights three conceptually distinct approaches that have been extensively applied in psychology.

Scoring motive imagery from TAT protocols. Murray's (1938) work on the TAT has had a profound effect on researchers interested in implicit aspects of human motivation. In a typical study, participants write brief stories about ambiguous black-and-white pictures. The essays are then scored for the presence of motive-relevant themes in participants' imagery. Whereas the original work by McClelland and Atkinson (1948) focused on how an aroused hunger motive surfaces in TAT fantasies, the main body of research has evolved around a small number of social motives. Various scoring systems are available for the need for achievement, the need for power, the need for affiliation, and the need for intimacy (for details, see Smith, 1992).

Recently, Winter (1994) integrated the different existing scoring systems into a unified manual that allows the simultaneous coding of achievement, power, and affiliation/intimacy imagery. According to this system, themes including improvement concerns such as "she wanted to find a better solution" are considered achievement imagery, whereas attempts to influence others (e.g., "he tried to convince him of the importance of this project") or references to status (e.g., "he impressed his friends with his new sports car") are interpreted as expressions of a need for power. Affiliation and intimacy themes are merged into one category and include both statements about friendships ("the two college friends were glad to see each other again") and intimate relationships ("they were young and in love"). A motive score is calculated by adding imagery scores across all stories and correcting for verbal productivity. More than 50 years after its development, TAT-based need assessment has recently experienced a surge in scientific attention (Schultheiss & Brunstein, 2001; Tuerlinckx, De Boeck, & Lens, 2002; Winter, John, Stewart, Klohnen, & Duncan, 1998).

Content Analysis of Verbatim Explanations. Peterson and Seligman developed Content Analysis of Verbatim Explanations (CAVE; Peterson, 1992) as a text analysis technique to complement questionnaire-based assessments of causal attributions. CAVE allows the scoring of any text document for the author's explanatory style.

The CAVE procedure involves two steps. First, all causal explanations in a text are identified. Trained scorers then rate each explanation on three dimensions (internality, stability, globality). Whereas "I can't go to the wedding because I have to go to a conference" is rated as not at all stable, "I didn't get the job because I am a woman" reflects a highly stable attribution. Similarly, "I did well on the paper because the assignment was easy" is considered highly external, whereas "I didn't get the job because I am too young" refers to a highly internal

TABLE 11.2

Overview of Nine Influential Quantitative Text Analysis Approaches in Psychology

Name	Reference	Linguistic parameters	Coding	Conceptual Classification			
				Aim	**Approach**	**Bandwidth**	**Focus**
Thematic Content Analysis	Smith (1992)	Need for power, need for achievement, need for affiliation, explanatory style, integrative complexity	judges	instrumental	thematic	specific	content/ style
General Inquirer	Stone et al. (1966)	Harvard III Psychosociological Dictionary, Stanford Political Dictionary, Need-Achievement Dictionary	computer	instrumental	thematic	broad/ specific	content/ style
Gottschalk–Gleser Method	Gottschalk et al. (1969)	Clinical phenomena (e.g., anxiety, hostility, social alienation, depression, cognitive impairment)	judges/ computer	instrumental	semantic	specific	content
Regressive Imagery Dictionary	Martindale (1990)	29 categories of primary process cognition (e.g., oral, sex, Icarian imagery), 7 categories of secondary process cognition (e.g., abstraction, social behavior), 7 categories of emotions (e.g., positive affect, anxiety)	computer	instrumental	thematic	broad	content
Analysis of Verbal Behavior	Weintraub (1981)	15 dimensions including pronouns (I, we, me), negatives (e.g., not, no, never), qualifiers (e.g., kind of), expressions of feelings (e.g., love), and adverbial intensifiers (e.g., really, so)	judges	instrumental	thematic	broad	style
TAS/C	Mergenthaler (1996)	Emotional tone, abstraction, referential activity	computer	instrumental	thematic	specific	style
DICTION	Hart (1984)	5 master variables (certainty, optimism, activity, realism, commonality) with a total of 35 linguistic subdimensions	computer	instrumental	thematic	broad	style
LIWC	Pennebaker et al. (2001)	82 variables; standard linguistic dimensions (e.g., pronouns, articles), psychological processes (e.g., emotion words, causation words), relativity (past tense, inclusive words), personal concerns (e.g., school, religion, sexuality)	computer	instrumental	thematic	broad	content/ style
LSA	Landauer et al. (1998)	n/a; 2 strategies with focus either on low (content approach) or high (style approach) frequency words	computer	representational	semantic	n/a	content/ style

cause (Peterson, Schulman, Castellon, & Seligman, 1992). Intensive coding training is offered.

The CAVE technique has been applied to a wide variety of text sources, including therapy protocols, newspaper articles, presidential addresses, personal letters, and TAT protocols. People's explanatory styles have been successfully linked to optimism, depression, and health behaviors (Peterson, 1992). The strength of the CAVE analysis lies in its theoretical foundation, its broad applicability, and its real-world relevance (Peterson, 1992).

Content analysis of conceptual/integrative complexity. Suedfeld, Tetlock, and their colleagues have developed a text analysis system to assess a person's information processing and decision making. Conceptual/integrative complexity (IC) measures the degree of differentiation and integration achieved in describing a phenomenon (Suedfeld, Tetlock, & Streufert, 1992).

Originally the Sentence/Paragraph Completion Test (S/PCT) was used as a source for assessing IC. In the S/PCT participants write open-ended answers to a series of sentence stems, such as "When I am criticized . . .," "When I don't know what to do . . .," or "When a friend acts differently. . . ." Each answer is then rated on a 7-point scale ranging from 1 (*no evidence of either differentiation or integration*) to 7 (*high differentiation and high integration*). In general, a high degree of differentiation is achieved when a phenomenon is acknowledged as having multiple causes and dimensions. Integration is obtained when interconnections are made between the acknowledged dimensions (Baker-Brown et al., 1992). IC scores are positively correlated with the total number of words in a text, the average sentence length, and the number of words with more than three syllables (Coren & Suedfeld, 1990).

Because the rating process involves subtle semantic inferences about the author's intention, intensive coder training is required (Suedfeld et al., 1992). More recently, IC analysis has been extended to the study of archival material. IC has been linked to a variety of social psychological topics such as attitude change, attribution, problem solving, and interpersonal communication (Suedfeld et al., 1992).

Summary and evaluation. Three influential thematic content analysis approaches have been reviewed. Several other coding systems are available but could not be included here (e.g., personal causation, deCharms, 1968; uncertainty orientation, Sorrentino, Roney, & Hanna, 1992; object relatedness, Rosenberg, Blatt, Oxman, McHugo, & Ford, 1994; for a more exhaustive review, see Smith, 1992). With regard to the four-dimensional conceptual framework, thematic content analysis is instrumental in its aim and thematic in its approach. It focuses either on verbal content (e.g., IC) or style (e.g., CAVE) and typically is specific in bandwidth. The fact that thematic content analysis involves human judges who make inferences about the meaning of a statement is typically considered a threat to its reliability. Generally, however, when quality standards such as appropriate test administration, careful judge training, and duplicate scoring of materials are met, good reliabilities are achieved (Schultheiss & Brunstein, 2001; Smith, 1992).

The main weakness of thematic content analysis lies in the time that judges spend coding verbal material. It has become increasingly attractive to replace moderately reliable and expensive human judges by perfectly reliable and cost-effective computer coders (cf. Hogenraad, 2003). Shapiro (1997) pointed to a weakness in this argument: Computer-based systems typically consist of two components, a processing device with the text analysis routine (e.g., the word count algorithm) and a dictionary with the linguistic information (e.g., lists of emotion- or achievement-related words). Whereas the processing device is 100% reliable, the deeper problem lies in the fact that coding ambiguity is shifted from the coding procedure to the construction of a comprehensive dictionary. Still, beyond their incomparable efficiency, computer codings also have the advantage of facilitating cross-study and cross-laboratory comparisons of findings.

The General Inquirer

In the early 1960s, Stone and his colleagues developed the "mother" of computerized text analysis, the General Inquirer (Stone et al., 1966). The General Inquirer is a compilation of a set of word count routines. It was designed as a multipurpose text analysis tool strongly influenced by both need-based and psychoanalytic traditions. Historically, three dictionaries, the Harvard III Psychosociological Dictionary, the Stanford Political Dictionary, and the Need-Achievement Dictionary have been applied the most with the General Inquirer. The Need-Achievement Dictionary was created to automate the judge-based scoring of TAT achievement imagery.

More important, the General Inquirer goes beyond counting words. In a two-step process, it first identifies so-called homographs (ambiguous words that have context-dependent meaning). It then applies a series of preprogrammed disambiguation rules aimed at clarifying their meanings in the text. For example, human judges score the statement "He is determined to win" as achievement imagery. The General Inquirer identifies the word *determined* as an ambiguous NEED word and *win* as an ambiguous COMPETE word (because they both can have nonachievement-related meaning) and codes a statement as achievement imagery only if both aspects are present and occur in the NEED–COMPETE order.

The General Inquirer is unique in its flexibility. It can be used to study virtually any topic of interest by creating user-defined dictionaries (e.g., Semin & Fiedler, 1988, 1991). Its most critical advantage, the power to perform context-dependent word counts, is also its most serious pragmatic drawback. The construction of a custom dictionary with the specification of disambiguation rules is time consuming and, in many cases, not well suited to the many ambiguous ways words are used. Nevertheless, the General Inquirer continues to shape the field of computerized text analysis. A third-generation version is now available for desktop computers as well as Internet usage. As shown in Table 11.2, the General Inquirer is instrumental in its aim and thematic in its approach. Its bandwidth and focus depend on the actual dictionary in use; the Need-Achievement dictionary, for example, is specific and content focused.

Gottschalk–Gleser Method of Content Analysis

Also in the 1960s, Gottschalk and his colleagues started developing what became known as the Gottschalk–Gleser Method of content analysis (Gottschalk, 1995). The Gottschalk–Gleser Method involves participants giving a 5-minute speech on a personal life experience. The verbatim transcripts are then submitted to a content analysis.

Several scales tapping into what Gottschalk calls "psychobiological dimensions" have been developed and validated. Most of the scales are derived from a psychoanalytic framework and are designed to diagnose clinical phenomena (Gottschalk, Stein, & Shapiro, 1997). Schizophrenic tendencies, for example, are meant to be revealed by the Social Alienation and Personal Disorganization Scale. Other scales diagnose depression, hostility, and cognitive impairment. Each scale consists of a number of subcategories that list the themes to be scored along with the respective scoring weights. The Anxiety Scale, for example, comprises death anxiety, castration anxiety, separation anxiety, guilt anxiety, and shame anxiety. Whenever one of these themes is mentioned, a weight is assigned according to the degree of (psychodynamic) association with the self (e.g., self: "I was scared I could die," +3 vs. other people: "He was scared he could die," +2 vs. objects: "The dog was scared it could die," +1).

The Gottschalk–Gleser Method relied originally on human judges. Recently, however, Gottschalk and Bechtel (1989, 1995) have introduced a computerized version. The computerized method is one of the few existing semantic text analysis tools in psychology (Popping, 2000). It uses a semantic grammar consisting of S-V-O templates to identify the action of the sentence (e.g., "to die") as well as the agent (e.g., "I" vs. "he") and—if applicable—the object. The Gottschalk–Gleser approach is specific in that it concentrates on selected clinical phenomena and focuses on the content of a person's statement.

Analysis of Artistic Change: Martindale's Regressive Imagery Dictionary

To identify regularities underlying changes in artistic work over time, Martindale (1990) developed a word count program that is based on the Regressive Imagery Dictionary. Martindale's (1990) theorizing starts from the observation that artistic work shows a steady increase in complexity over time. He explains this increase by drawing on two fundamental psychological processes: humans' preference for medium levels of arousal (and hence moderately complex sensory input) and the physiological mechanism of stimulus habituation (leading to changes in what is considered moderately complex). Grounded in psychodynamic thinking, he plotted how two major linguistic dimensions in literature, primordial (i.e., primary process) and conceptual (i.e., secondary process) cognition, have changed over the decades.

Martindale's Regressive Imagery Dictionary has been translated into several languages (e.g., French, German, and Portuguese). The English version is composed of about 3,200 words and word stems that fall into 29 categories of primary process cognition (e.g., regressive cognition, Icarian imagery), 7 categories of secondary process cognition (e.g., abstraction, social behavior), and 7 emotion categories (e.g., positive affect, anxiety).

Over the last 30 years, Martindale (1990) has accumulated an impressive body of studies that identify linguistic indicators of an aesthetic evolution. Unfortunately, his work has not enjoyed widespread attention in mainstream psychology (cf. Bestgen, 1994; Hogenraad, McKenzie, Morval, & Ducharme, 1995). As depicted in Table 11.2, Martindale's text analysis approach is instrumental in aim, thematic in approach, and broad in bandwidth. It focuses on the content of literature from a psychodynamic perspective.

Weintraub's Analysis of Verbal Behavior

Weintraub's (1981, 1989) work on verbal mannerisms was inspired by the clinical observation that individuals speaking under stress often reveal important information about their psychological adjustment. Drawing on his medical training and practice, Weintraub argued that psychological defense mechanisms manifest themselves in speech patterns obtained under mildly stressful conditions. He assessed these defense mechanisms from the language that participants spontaneously use when they talk for 10 minutes about a personal topic (Weintraub, 1981).

Unlike most other word count approaches, Weintraub's linguistic analysis is performed by naïve judges who "can score . . . [the transcripts] without extensive knowledge of lexical meaning" (Weintraub, 1989, p. 11). The linguistic parameters that he is interested in are largely intuitively derived and drawn from his clinical experiences. Weintraub's most recent work has focused on 15 linguistic dimensions, including three pronoun categories (I, we, me), negatives (e.g., not, no, never), qualifiers (kind of, what you might call), expressions of feelings (e.g., I love, we were disgusted), and adverbial intensifiers (really, so).

Weintraub has explored verbal behavior in multiple ways. In addition to his main interest, the language of psychopathology, he also analyzed the Watergate transcripts, characterized speaking styles of post–World War II U.S. presidents, identified linguistic correlates of intimacy, and related language use to personality. Weintraub's analyses are instrumental in aim, are thematic in approach, capture a broad spectrum of language use, and are stylistic in focus (see Table 11.2).

Analyzing Emotion–Abstraction Patterns: TAS/C

Mergenthaler and his research group use text analysis to characterize key moments in psychotherapy sessions. They developed a computer program called TAS/C that focuses on two language dimensions—emotional tone and abstraction. According to Mergenthaler's theory, emotion–abstraction patterns occur periodically in psychotherapy sessions with insight processes (abstraction) following emotional events (emotion) with a time lag (Mergenthaler, 1996).

For the analysis of emotional tone, defined as the density (rather than the valence) of emotion words, a dictionary with more than 2,000 entries was developed. The final list of emotion words comprises three dimensions (pleasure, approval, and attachment) and captures roughly 5% of the words of a text (Mergenthaler, 1996). Abstraction is defined as the number of abstract nouns in a text. Abstract nouns are identified via suffixes such as -ity, -ness, -ment, -ing, or -ion. The abstraction dictionary includes 3,900 entries and captures about 4% of the words of a text.

TAS/C analysis of emotion–abstraction patterns has been applied to verbatim therapy protocols (Mergenthaler, 1996) and attachment interviews

(Buchheim & Mergenthaler, 2000). More recently, TAS/C has been extended to include a measure of referential activity. Referential activity refers to the ability to verbalize nonverbal experiences and is characterized in speech by concreteness, specificity, clarity, and imagery (Mergenthaler & Bucci, 1999). The TAS/C approach is instrumental in its aim, is thematic in its approach, and concentrates on two specific stylistic aspects of language use in psychotherapeutic settings.

Analyzing Verbal Tone With DICTION

Hart (1984, 2000) is interested in word choice in political communication. Over the last two decades he has developed a computerized word count program called DICTION (Hart, 2001). DICTION is designed to reveal the verbal tone of political statements by characterizing text on five statistically independent master variables: activity, optimism, certainty, realism, and commonality. The rationale behind these master variables is that "if only five questions could be asked of a given passage, these five would provide the most robust understanding" (Hart, 2001, p. 45). The five master variables are composed of 35 linguistic subfeatures (e.g., optimism is composed of the subfeatures praise, satisfaction, inspiration, blame, hardship, denial).

DICTION relies on 10,000 search words that are assigned to the categories without overlap. The output is either a profile of absolute values or norm scores that is based on 20,000 samples of verbal discourse. Special features of DICTION are the ability to learn, that is, to update its database with every processed text, and a statistical weighting procedure for homographs. DICTION has been used to analyze presidential and campaign speeches, political advertising, public debates, and media coverage. It is instrumental in aim, is thematic in the approach, captures language at a broad level, and focuses on stylistic aspects of texts.

Linguistic Inquiry and Word Count

Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2001) was originally developed in the context of Pennebaker's work on emotional writing. It was designed to reveal aspects of writing about negative life experiences that predict subsequent health improvements (Pennebaker & Francis, 1996; Pennebaker, Mayne, & Francis, 1997). More recently LIWC has been used to analyze language use in a wide variety of text sources including literature, personal narratives, press conferences, and transcripts of everyday conversations (Pennebaker et al., 2003).

LIWC searches for over 2,300 words or word stems within any given text file. Independent judges previously categorized the search words into 82 language dimensions. These dimensions include standard linguistic categories (e.g., articles, prepositions, pronouns), psychological processes (e.g., positive and negative emotion words, words referring to cognitive or social processes), relativity-related words (e.g., time, motion, space), and traditional content dimensions (e.g., sex, death, job). Most LIWC dimensions are hierarchically organized; for example, the word *cried* falls into the four categories of sadness, negative emotion, overall affect, and past-tense verb. The program also offers the option to create user-defined categories.

Although some LIWC dimensions are based on specific psychological theories (e.g., inhibition words, discrepancy words), most categories extract information at a basic grammatical (e.g., pronouns, articles, prepositions) and psychological level (e.g., emotion words). LIWC is instrumental in its aim and thematic in its approach. It captures broad aspects of language use. Currently, LIWC has been found to be most effective in tracking stylistic aspects of language use. However, with its traditional content categories, it also allows for a basic analysis of text content (e.g., achievement, religion, sexuality). Recently, Spanish, German, and Italian versions of the LIWC dictionary have been developed and tested for equivalence to the original English version. LIWC has been applied to a wide spectrum of research questions in social, personality, and clinical psychology, including coping with trauma, depression, suicidality, gender differences, personality expression, and aging (Groom & Pennebaker, 2002; Pennebaker et al., 2003).

Extracting Word Patterns: Latent Semantic Analysis

Latent Semantic Analysis (LSA; Foltz et al., 1998; Landauer & Dumais, 1997; Landauer, Foltz, &

Laham, 1998) is a semantic text analysis strategy and concerned with the use of words in their context. Compared to most existing semantic text analysis programs, however, LSA does not adopt the top-down strategy of specifying a semantic grammar and looking at the occurrence of S-V-O constellations. Instead—in a bottom-up manner—it distills information about the semantic similarity of words by analyzing their usage across a large body of text.

Applying singular value decomposition, a mathematical data reduction technique akin to factor analysis, LSA creates a multidimensional semantic space that allows one to calculate the similarity between any two words used in a given body of text by comparing their coordinates in the semantic space. If, for example, the words *patient* and *physician* consistently co-occur in a sentence across a large amount of text, LSA assigns them similar factor weights. Ignoring syntactical information, LSA infers similarity in meaning from patterns of word co-occurrences. LSA was initially developed as a search engine with a focus on words that carry content (i.e., nouns, verbs, adjectives). This has lead to its application as a tool to measure textual coherence (e.g., Foltz et al., 1998) and to provide computerized tutoring (e.g., Graesser et al., 1999).

More recently, LSA has been adapted to analyze textual style. For this, LSA ignores low-frequency content words and focuses on high-frequency words that have minimal semantic function (i.e., pronouns, articles, prepositions). In a reanalysis of three studies on the salutary effects of emotional writing, Campbell and Pennebaker (2003) linked an LSA measure of similarity in people's essays across 3 days of writing to their subsequent health. They found that similarity in the use of common words, especially personal pronouns, was negatively related to health benefits. This study underscores that LSA is not an esoteric tool for cognitive scientists, but can offer a fresh perspective on persistent problems in social psychology.

Clearly, LSA's word pattern analysis has limitations (Perfetti, 1998). Its inability to consider syntactic structure or to make use of acquired word knowledge certainly distinguishes it from human coders. However, Landauer et al. (1998) argued that "one might consider LSA's maximal knowledge of the world to be analogous to a well-read nun's knowledge of sex, a level of knowledge often deemed a sufficient basis for advising the young" (p. 261). LSA is representational in its aim and semantic in the approach. As explained earlier, it can focus on low-frequency words that carry content or on high-frequency words that convey linguistic style.

Summary and Evaluation

This section reviewed nine influential text analysis strategies in psychology. The selected approaches span a broad spectrum of methodological and theoretical orientations. How should a researcher decide which one to use? The most immediate question is whether the options are restricted to computerized solutions or whether the burden of manual coding appears tolerable (Smith, 1992; Weintraub, 1981). Another question concerns what kind of analysis a researcher is interested in. The four-dimensional framework was introduced to help with this question.

Over and beyond this, however, other characteristics of the programs also help determine the most appropriate solution for a given research project. Several of the reviewed approaches emerge from psychodynamic theorizing. For researchers whose interest lies in this area, the solutions offered by Gottschalk (1995), Martindale (1990), Weintraub (1981), or Mergenthaler (1996) are good choices—with the Gottschalk–Gleser Method having the strongest clinical focus, Martindale's Regressive Imagery Dictionary being particularly useful for the analysis of literature, and Mergenthaler's TAS/C being the ideal tool for the analysis of therapy protocols. DICTION (Hart, 1984) assesses psychological variables at a comparatively abstract level and—because of its background in communication research—seems most useful for the study of political communication and persuasion. For researchers interested in basic grammatical text features (e.g., pronouns, articles, prepositions) or low-level psychological constructs (e.g., emotional, cognitive, or social processes), LIWC (Pennebaker et al., 2001) offers an extensively validated solution. The General Inquirer (Stone et al., 1966) also captures a wide variety of psychological parameters and, in its most recent version, includes an operationalization of Semin and Fiedler's (1988, 1991) Linguistic Category Model. Finally, LSA (Landauer et al., 1998) is a powerful text analysis tool that is not word count based and has applications in modeling cognitive processes such as knowledge representation, coherence, and perspective taking.

QUANTITATIVE TEXT ANALYSIS: A METHOD REFLECTION

The final section of this chapter steps back and reflects more broadly on the potentials and pitfalls of text analysis as a scientific method. The discussion revolves around three major questions: The first question asks what makes text analysis an attractive method for psychology. The second question looks at text analysis from a measurement perspective and asks to what extent is verbal data psychometrically good data. The third question is fueled by the apparent paradox that on the one hand, the vast majority of existing text analysis programs are word count based but that, on the other hand, simple word count solutions often appear overly simplistic and fraught with problems. How far can we go with simply counting words?

What Makes Text Analysis an Attractive Method for Psychology?

From the time we get up in the morning—listening to the radio or reading the newspaper—until we go to bed—watching TV or reading a book—we are surrounded by words. Every day we have dozens of conversations, make numerous phone calls, write and receive an increasing number of e-mails, surf the Internet, and chat in chat rooms. As teachers we assign writing assignments and grade essays. As researchers we use language to communicate with our participants; we collect responses to open-ended questionnaires, conduct interviews, videotape discussions, and record conversations. It is overwhelming how our daily lives are saturated with words. Thus, it is surprising how little psychologists have used language as a source of data.

With the advent of the Internet, various new opportunities for studying linguistic phenomena have opened up. Without running a single participant, researchers can now collect large amounts of text from personal Web pages, chat room conversations, message board entries, and e-mails (e.g., Cohn, Mehl, & Pennebaker, in press). Also, all major newspapers, magazines, periodicals, and journals are now available online and maintain comprehensive electronic archives. Important statements of public figures such as presidential addresses or press conferences are usually available soon after they occur—often already in transcribed form. Virtually any song's lyrics and even entire movie scripts can be downloaded from the Web. In short, text analysis researchers never experience a data shortage.

However, there is more to text analysis than the opportunity to draw on easily available data. As a method for analyzing archival data it offers another critical advantage (Lee & Peterson, 1997; Simonton, 2003; Winter, 1992). The data collection is less constrained than in most other methods. Survey studies yield scaled answers on a limited set of items—selected by the investigator on conceptual grounds prior to the onset of the study. Questionnaires work by a "what you ask is what you get" principle. No further information can be obtained once the data are collected. Open-ended questions, essays, or other verbal productions are different; they allow researchers to go back to the data and explore aspects that one had not originally considered.

Going back to our initial example about students' motivation to seek out a doctor, for instance, one might later become interested in whether self-focused attention operationalized as the use of first-person singular ("I") could predict who goes to the doctor. The data is also available for unrelated research questions such as sex differences in language use (Groom, Stone, Newman, & Pennebaker, 2004). It is even possible for other researchers now or in the future to analyze the data using their own text analysis approach and interpretative framework. The analysis of verbal material provides a flexibility that is hard to obtain with other methods.

So far, the vast majority of text analysis researchers have relied on a single type of text source. From a multimethod perspective, for a more elaborate understanding of how people use language it is necessary to start comparing language effects across text sources, genres, or contexts. For example, are there systematic differences in the way humans express themselves in written as compared to spoken language (Biber, 1988; Mehl & Pennebaker, 2003; Weintraub, 1981)? Or is language use in e-mails more similar to how people actually talk or write letters (Baron, 1998)? Identifying the degree of linguistic convergence and uniqueness across different language sources is an important area for future research (Pennebaker et al., 2003).

Is Verbal Data Psychometrically Good Data?

There might be many good reasons to use text data for psychological research. From a measurement perspective, one of the most important questions concerns the extent to which verbal data is psychometrically good data. Unfortunately, it is common for text analysis researchers, after developing a new method, to proceed to its application without establishing its psychometric properties. Thorough construct validation in the area of text analysis is yet rare. However, at least two notable exemptions to this rule deserve to be mentioned. A large body of research has established the validity of TAT-based motive measures. From this it has become clear that implicit motives (a) can be reliably assessed with the TAT (Lundy, 1988; Smith, 1992; Tuerlinckx et al., 2002; Winter & Stewart, 1977), (b) are distinct from self-reported motives and traits (King, 1995; Schultheiss & Brunstein, 2001), and (c) uniquely predict types of behavior (McClelland, Koestner, & Weinberger, 1989; Winter et al., 1998).

The basic psychometric properties are also comparatively well understood for word count–based measures. Across a series of studies, the words that people use in their spoken and written language have emerged as stable over time and across context (Gleser, Gottschalk, & Watkins, 1959; Mehl & Pennebaker, 2003; Pennebaker & King, 1999; Schnurr, Rosenberg, & Oxman, 1986). Also, spontaneous word choice shows reliable and theoretically meaningful associations with demographic variables (Groom et al., 2004; Pennebaker & Stone, 2003) and traditional personality measures (Pennebaker et al., 2003), but also predicts, for example, real-life health behaviors over and beyond the Big Five dimensions (Pennebaker & King, 1999).

To summarize, the question to what extent text analysis yields good data from a measurement perspective is important and needs to be answered for each method separately. So far, at least for TAT-assessed motives and word count–based measures, the existing evidence suggests good psychometric properties. Thorough construct validation that, for example, establishes aspects of convergent validity between different text analysis methods (e.g., emotion words across different programs) and between text analysis methods and other psychological methods (e.g., self-reported, observed, and linguistic measures of emotions) are needed.

How Far Can We Go With Counting Words?

Given that word count–based measures possess rather good psychometric properties, how far can we go with counting words? Frequently researchers voice their scientific disdain for text analysis programs that are unable to distinguish between sentences as simple as "the dog bit the man" and "the man bit the dog" (Hart, 2001, p. 53). Its blindness to context makes word-count approaches sometimes appear painfully dumb. Not only are they unable to pick up irony or sarcasm (e.g., "Thanks a lot," accompanied by a roll of the eyes) and metaphoric language use (e.g., "He had the key to her heart"), but they also confuse words that have different meanings in different contexts (e.g., "What he did made me *mad*" vs. "I'm *mad* about the cute person in my class"). In a discussion of the shortcomings of a program such as the General Inquirer, Zeldow and McAdams (1993) went as far as to entirely question the value of lower-level word counts.

Over the last five decades, however, word-count approaches have repeatedly demonstrated their potentials in virtually all domains of psychology (e.g., Gottschalk, 1995; Hart, 1984; Martindale, 1990; Pennebaker et al., 2003; Stone et al., 1966; Weintraub, 1981). Often, to test psychological hypotheses, it is not necessary to specify grammatical relationships between themes; instead, it is sufficient to know *that* certain themes (co-)occur in a text. In fact, Hart (2001) even construed thematic text analysis' blindness toward context as its biggest advantage. Because humans so readily understand the communicative meaning of words, having a computer that counts themes under full neglect of their semantic surroundings provides researchers with information that is largely inaccessible to self-report or observational methods.

If one accepts that the study of words can be psychologically meaningful, which words should researchers focus on? It is interesting that virtually every text analysis approach has started from the assumption that emotional states can be detected by studying the use of emotion words (cf. Bestgen, 1994). The reality is that in daily speech, emotional writing, and even affect-laden poetry, less than 5% of the words can be classified as emotional (Mehl & Pennebaker, 2003; Pennebaker & King, 1999). From an evolutionary perspective, it is unlikely that language has evolved as a vehicle to express emotion. Instead, humans use intonation, facial expression, or other nonverbal cues to convey feelings. Emotional tone is also expressed through metaphor and other means not related to emotion words. Taken together, embarking on emotion words to study human emotions has not emerged as a particularly promising strategy (Pennebaker et al., 2003).

Content-based dictionaries are generally comprised of word categories that the researcher created based on more or less empirically supported intuitions of what words are indicative of certain themes (e.g., the word *football* is indicative of the theme *sport*). Hence, content dictionaries always have a subjective and culture-bound component (Shapiro, 1997). Markers of linguistic style, however, are generally associated with relatively common "content-free" words, such as pronouns, articles, prepositions, conjunctives, and auxiliary words—also referred to as particles (Miller, 1995). Particles are easier to handle because their meaning is less ambiguous, less context bound, and more determined by grammatical rules. In the English language, there are fewer than 200 commonly used particles, yet they account for over half the words we use.

From a psychological perspective, not all particles are equal; personal pronouns have emerged as particularly revealing (Pennebaker et al., 2003). Although the use of the first-person singular ("I"), for example, indicates an explicit distinction that speakers make between themselves and their social world, the use of the first-person plural ("we") suggests speakers experience themselves as part of a larger social unit. Empirically, the use of the first-person singular is associated with age, sex, neuroticism, depression, illness, and more broadly, attention focused on the self (Pennebaker et al., 2003). The use of second-person ("you") and third-person ("he," "she") pronouns, by definition, show that the speaker is socially engaged or aware. So, it becomes clear that in the conversational context, pronouns have important social implications. The empirical evidence to date underlines this by pointing to their role as powerful markers of psychological processes and predictors of mental and physical health (Pennebaker et al., 2003).

SUMMARY

One purpose of this chapter was to demonstrate that quantitative text analysis is a powerful, efficient, and easy-to-use tool for psychological research. The review of nine different text analysis strategies showed that the spectrum of existing applications is wide—although some methods continue to rely on human judges, the majority use computers to count isolated words, and a few harness more sophisticated techniques to assess the semantic relationships.

Where will the field go from here? Extrapolating from current progress in artificial intelligence, there is no doubt that in the years to come, text analysis applications will become increasingly complex (West, 2001). Will simple word count programs soon be declared scientific history? Considering that they are currently the only solutions in which complete automation has been achieved (Shapiro, 1997), this scenario seems unlikely. With their ability to process large amounts of texts in a matter of seconds without any preformatting, word-count programs have a tremendous pragmatic advantage over more sophisticated tools that require a human labor force for extensive data preparation and text parsing.

Hence, researchers who are interested in text analysis are encouraged to be aware of the "bigger is better" fallacy. Tempting as it might seem, the assumption that more technically advanced programs will necessarily be more appropriate for addressing a researcher's question does not always hold up. Simple word-count approaches—crude, fuzzy, and error prone as they are—can often go a long way. After all, by only using a simple home-made telescope and not high-resolution satellite pictures, Galileo was able to detect the four moons of Jupiter.