Published in final edited form as:

J Nerv Ment Dis. 2012 August ; 200(8): 668–675. doi:10.1097/NMD.0b013e3182614127.

Textual Data in Psychiatry:

Reasoning by Analogy to Quantitative Principles

Suzanne Yang, MD^{*,†}, Edward P. Mulvey, PhD[†], and Bruno Falissard, MD, PhD[‡]

^{*}Mental Illness Research, Education and Clinical Center (MIRECC), Veterans Administration Pittsburgh Healthcare System, Pennsylvania

[†]Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, Pennsylvania

[‡]Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France

Abstract

Personal meaning in subjective experience is a key element in the treatment of persons with mental disorders. Open-response speech samples would appear to be suitable for studying this type of subjective experience, but there are still important challenges in using language as data. Scientific principles involved in sample size calculation, validity, and reliability may be applicable, by analogy, to data collected in the form of words. We describe a rationale for including computer-assisted techniques as one step of a qualitative analysis procedure that includes manual reading. Clarification of a framework for including language as data in psychiatric research may allow us to more effectively bridge biological and psychometric research with clinical practice, a setting where the patient's clinical "data" are, in large part, conveyed in words.

Keywords

Language; meaning; research methods; subjective experience; textual data

THE RELEVANCE OF SUBJECTIVE MEANING

The attribution of subjective meaning to events in daily life is so pervasive that it would seem to defy scientific investigation. This type of subjectivity is at once universal and highly specific to each individual, and it is situated in the context of social relationships and personal history. The very idea of finding a rational approach to study subjective meaning is daunting.

However, the patient's subjective point of view is central to psychiatric practice (Kendler, 2005). Mental health professionals are constantly faced with patients who have firmly held beliefs that no one else finds plausible, profound sadness that is incommunicable to others, or a sense of desperate isolation. The effort to comprehend a patient's experience of having a mental illness is a key element of thoughtful and empathic intervention, even if the clinician finds that, ultimately, she cannot fully know the patient's world. In this context,

Copyright © 2012 by Lippincott Williams & Wilkins

Send reprint requests to Suzanne Yang, MD, Law and Psychiatry Program, University of Pittsburgh and the Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213. yangs@upmc.edu.

DISCLOSURES

The authors declare no conflict of interest.

language may be a last vestige of relatedness, even when its content expresses subjective experience that is colored or substantially distorted by symptoms of a mental disorder.

The predominant reaction of researchers to the challenge of understanding mental disorders, however, has not been to study the disorders from the perspective of how the individual attributes meaning to specific experiences. Scientific inquiry into mental disorders is instead largely based on developing ways to measure reported symptoms, functioning, and outcomes, and examining the quantitative relationships among these constructs. These techniques are able to index subjectivity (Falissard, 2001), but not in terms of the person's point of view, as it is situated within life circumstances and their meaning. As a result, we know how to scale many aspects of mental functioning, symptoms and behavior, and we have made significant therapeutic advances, but we actually know very little about what it is like for individuals to experience these phenomena.

The study of subjective experience through words and their meaning provides a unique mode of access to the mental life of patients. The complex task of investigating the subjective aspects of mental disorders requires the further development of language-based methodologies designed to characterize the world in the patient's own terms and to situate such findings in relation to existing knowledge. In clinical practice, the patient's subjective experience is conceptualized idiographically, in terms of meanings that are highly personal and specific to the case, given that the therapeutic process is individualized and must make sense within the world as constructed by the patient. In psychiatric research, however, many questions can be answered only within large populations, nomothetically across subjects. In this latter type of endeavor, individual differences in subjective experience can still be valuable to consider, if there are sufficiently clear procedures for using language as data across individuals, with mixed-methods analyses to assess their effects. Doing so may enhance the value of clinical studies by accounting for potential sources of "noise" and by construing them instead as meaningful information (Hohmann and Shear, 2002).

Investigators have long posited that language behavior is a relevant indicator in studies of psychopathology (Andreasen and Pfohl, 1976; Chevrie-Muller et al., 1985; Newman and Mather, 1938; Weintraub and Aronson, 1962). A substantial number of studies have examined linguistic features of disorders such as schizophrenia, where language dysfunction is prominent (Covington et al., 2005; Kuperberg et al., 2010). These studies have clarified language structure and function within a quantitative framework (Hoffman et al., 1986; Oxman et al., 1988), and some studies have effectively used computer-assisted linguistic techniques to predict neuropsychiatric characteristics (Gottschalk et al., 1997). Other investigators have examined subjective perceptions of psychiatric illness using qualitative analysis and manual reading or coding (Jenkins and Carpenter-Song, 2009; Roe et al., 2008; Sajatovic et al., 2009). Our focus in this article was on the development of a framework and principles for combining such techniques to optimize the ways that subjective meaning in mental disorders is examined in research studies.

We first discuss the value of conducting research on the subjective experience of mental illness using data that are collected in the form of words and briefly review the existing range of approaches for doing so. We then examine more specifically the use of brief speech samples as units of data that are flexible enough to provide personalized depictions of inner states and yet structured enough to provide a consistent basis for analysis. Although the content may be subjective, data collected in the form of speech, once transcribed, provide a concrete substrate for scientific analysis. We describe a previous study as an illustration of the types of challenges that are commonly encountered and use this study as an example to critically examine solutions available for systematically accessing subjective experience across individuals.

THE PLACE OF SPEECH AS TEXTUAL DATA IN PSYCHIATRIC RESEARCH

Contemporary research questions in psychiatry are often posed in terms that favor their examination through statistical methods. Although textual data are sometimes collected in the form of open-response questions or in focus groups, they are often used in an ancillary or “behind the scenes” manner to support the refinement of research questions and survey items. The exact procedures for collecting, collating, and analyzing such information are seldom reported. This type of data, however, need not be regarded as merely exploratory, in the shadow of a “real study” with numerical data.

Although a substantial body of qualitative research has made effective use of textual data within existing psychiatric research paradigms, this approach has still been relatively underused in research programs addressing the most salient questions that face mental health practitioners (Hohmann and Shear, 2002). “Qualitative” research, using methods adopted from social science, refers to a vast and diverse domain with distinct values and procedures. These approaches share a supposition that the investigator’s own subjectivity is an instrument of analysis deployed through procedures carefully delimited by reflexive awareness of potential biases and assumptions (Alvesson and Skoldberg, 2009; Maxwell, 2005). The investigator’s beliefs and world-view guide and enrich the analysis, through critical examination of the theoretical lens that led to the research question. These theoretical foundations are traditionally grouped into schools of thought with differing ontologies: a) *post-positivist or empirical* approaches, such as grounded theory (Glaser and Strauss, 1967), which suppose the existence of externally verifiable truths and realities that are accessible through systematic study; b) *constructivist or hermeneutic* approaches, which rely upon the progressive development of interpretations relative to the subjective attitudes of the observer and observed parties; c) *critical theory*, which views science as embedded in political-ideological conditions and seeks to expose, challenge, or modify these; and d) *post-structuralist* strategies, which emphasize the inherent instability of constructs to undermine authority claims and epistemological certainties (Alvesson and Skoldberg, 2009; Hatch, 2002). Although a detailed discussion of these different theoretical orientations is beyond the scope of the present article, awareness of the suppositions and ontological preferences that guide formulation of a research question is useful for emphasizing the range of available approaches, within which quantitative reasoning represents a specific set of values. Methods can therefore be viewed as technical representations of epistemological or ontological positions, and methodological choices therefore define the range and types of knowledge that research can produce (Bryman, 1984/2003).

Qualitative approaches have been applied to psychiatry in a number of incisive ways. These have included analyses that reveal larger contextual issues in psychiatry, such as the ways that societal views of gender have shaped concepts in psychopathology (Ehrenreich and English, 1978; Lunbeck, 1994; Showalter, 1985). A longstanding historiographic tradition has brought insights regarding psychiatric practice as it has evolved over centuries and increased awareness of how scientific values are embedded in the social, cultural, and perhaps political frameworks in which they are generated. Michel Foucault’s critique of the social control function of psychiatry, for instance, stimulated discussion of how practitioners are inherently situated within institutions and systems that reinforce existing power relations (Foucault, 1961/2006; see also Castel, 1988; Goldstein, 2002). Gauchet and Swain (1980/1999), in their pointed response to this work, presented evidence that psychiatric confinement served the principal aim of providing treatment and empathic assistance, despite the undeniable social control function of the asylum. More recent work has examined the fluid boundary between the normal and the pathological, using social phobia and depression as examples (Horwitz and Wakefield, 2007; Lane, 2007). Historical methods, using adaptations of ethnographic approaches, have also provided crucial insights

into the contexts within which psychiatric knowledge is produced and how it changes according to new needs in different eras (Horwitz, 2002). Perhaps more so in psychiatry than in other medical sciences, inquiries such as these can reveal the potential relevance of subjective experience, situated in its social context, in the exploration of pathology.

Individual case studies (Widlöcher, 1999) and ethnography, both well-developed practices based on hermeneutic principles, have also enriched psychiatric thinking. In-depth ethnographic analysis through participant observation has been used to describe psychological aspects of imprisonment (Toch, 1975, 1996; Toch et al., 2002), the subjective perspectives of persons with schizophrenia and their treatment teams (Barrett, 1996; Estroff, 1981; Jodelet and Duveen, 1991), informed consent processes in psychiatry (Lidz et al., 1984), as well as psychiatric education (Luhmann, 2000). More recently, researchers have effectively deployed qualitative and mixed methods within clinical studies of psychiatric patients (Cusack et al., 2007; Robins et al., 2005).

Qualitative approaches are often described as idiographic, or local and case specific, but the notion that qualitative work is always distinct from nomothetically-based inquiry may be simplistic. Recent statistical research demonstrates that idiographic reasoning can also be expressed quantitatively (Haynes et al., 2009), whereas qualitative work can be nomothetic and seek to produce knowledge regarding large-scale phenomena or even aim for generalizability (Donmoyer, 1990/2000; Gomm et al., 2000; Schofield, 1990/2000). The idiographic-nomothetic distinction today calls for further clarification, more precise definition, or possible dismantling (Krauss, 2008).

Choices in strategy for handling textual data can serve as an opportunity to explore the potential synergy between classic qualitative and nomothetic approaches. Data collection via open-response questions may be particularly useful because it allows the subject to shape the response, in contrast to standardized rating scales where the theoretical framing of questions has been delimited in advance. In disadvantaged and stigmatized groups, such as mentally ill patients or prisoners, open-response conditions can bypass the defensive attitude that subjects may have toward highly structured rating instruments, thereby providing more nuanced and perhaps more accurate data about their subjective experience.

We came to appreciate the value and potential of this approach in a previously reported qualitative study on the impact of incarceration in mentally ill and non-mentally ill prisoners, which used transcribed speech samples (Yang et al., 2009). We found, surprisingly, that far from being reticent or “hardened,” these incarcerated individuals were willing to reveal their inner lives and reflections, albeit briefly, by speaking to interviewers with whom they had time to develop only minimal rapport. This resulted in a richly textured data set of verbal responses across a group of 59 subjects.

In the course of this work, we conducted a literature search on “qualitative” methods used by other researchers to handle this type of data. This yielded two main types of materials: a) articles and books outlining general principles for rigor in qualitative studies (*e.g.*, Glaser and Strauss, 1967; Mays and Pope, 2000; Pope et al., 2000; see also Altheide and Johnson, 1998; Corbin and Strauss, 2008; Creswell, 2007; Hatch, 2002; Maxwell, 2005) and b) specific research articles that presented the methods and results of their study in a detailed manner (*e.g.*, Lim et al., 2004; Stamm et al., 2005).

The articles outlining general principles, although helpful, provided little concrete guidance regarding how to choose particular approaches for data collection and analysis. On the other hand, research articles were useful in providing examples of analytic procedures but rarely explained the reasoning behind choices regarding data collection methods, data structure, or analysis. Qualitative approaches are often crafted in relation to each situation anew, in

accordance with the specific needs of the study at hand (Bowen, 2008). It is not surprising, therefore, that the choice of methods does not lend itself to explicit directives.

This gap in the literature is more than just the product of a need for flexibility in designing qualitative approaches. It is also more broadly symptomatic of the relative neglect of strategies using linguistic data in the field of psychiatry, despite the existence of viable technologies (Pennebaker et al., 2003). By clarifying the scientific principles and framework for using such data, research can be designed to render the use of this type of data more accessible and relevant to researchers for understanding the subjective experience of mental illness.

Here, we focus on applying the analysis of language within a mixed-methods framework involving brief speech samples obtained systematically from multiple subjects. Although the growing literature on mixed methods has become increasingly precise in specifying methodological principles (Dellinger and Leech, 2007; Johnson et al., 2007; Teddlie and Yu, 2007), we emphasize an area that still requires further development regarding the analysis of linguistic data.

STUDY OVERVIEW: LONG-TERM INCARCERATION AND THE IMPACT OF MENTAL ILLNESS

A previous study (Yang et al., 2009) illustrates some of the principles and choices involved in using textual data. As part of a large-scale clinical epidemiological survey of the psychological health of inmates, 59 prisoners who had been incarcerated for 10 years or longer gave their accounts of the subjective impact of imprisonment. A brief time-limited interview using three standard questions with open responses (*e.g.*, “How do you think that prison has modified your psychological state?”) was used. After each question, the prisoner was given 30 seconds to gather his thoughts before speaking freely for up to 2 minutes in reply. Data from structured diagnostic instruments were also obtained.

The transcribed responses were analyzed using two complementary methods. First, a primary reader conducted a thematic qualitative review of the responses, organizing key words, exemplary phrases, and themes into a preliminary model. A second reader reviewed the transcripts independently and also identified major categories. A revised model was developed after discussion by these two investigators. Second, the combined text of the subject responses was examined using a computer-assisted linguistic analysis software package (Image, 2007). The ALCESTE program generated groupings of frequently associated words, and these were compared with the content-based categories derived from the qualitative review of the text, further refining and differentiating these categories.

Finally, using results from a structured diagnostic procedure, subgroups of severely mentally ill prisoners and prisoners with no psychiatric disorder were identified and their responses were compared, using both qualitative reading and a feature of the ALCESTE program that identifies word use specific to predefined sets of cases. Many prisoners spoke introspectively about topics such as the purposes of their sentence, the passage of time, and difficulties in self-expression. In contrast, when analyzed separately, the subgroup of prisoners with a severe mental illness expressed notably more hostile and persecutory interpretations of the prison environment, suggesting that the psychological impact of incarceration in this group was strongly influenced by the presence of psychiatric symptoms.

The analytic procedures used illustrate several points regarding the relevance and feasibility of using linguistic data to address research questions involving subjective meaning. Two main sets of design issues, a) sampling and data format and b) validity and reliability, are

discussed below, drawing upon this study to suggest general principles in using this type of data.

Sampling and Data Format

Sampling is an essential design feature that limits the types of conclusions that are reasonable to draw, regardless of whether the data are linguistic, numerical, or categorical. Determination of an appropriate sample size is also a critical factor when designing a study or applying for research funding. Unlike the study described here, traditional qualitative data collection involves open interviews of varying duration, using sampling strategies that are *purposive* with the aim of testing elements of an emerging theory (Corbin and Strauss, 2008; Glaser and Strauss, 1967). In such a context, there is no numerical procedure for determining “adequate” sample size, with the standard guideline being that of reaching “theoretical saturation”—the point where no new categories are generated upon the collection of further data—rather than a specified number of subjects (Bowen, 2008; Creswell, 2007; Onwuegbuzie and Leech, 2007). Mixed-methods studies often use the number of subjects required by the quantitative component as a guide for the qualitative sampling. In principle, however, the situation can arise where vastly different sample sizes are needed for the two components of a given study (Creswell and Plano Clark, 2007).

There are, however, principles that effectively guide the determination of sample size when collecting relatively brief speech samples in a defined study population. Estimation of appropriate sample size in studies involving linguistic data requires a balancing of the *number* and type of primary subjects with the *amount of data* collected of each subject. Consideration of these issues in light of the research question at hand allows for a reasoned determination of the appropriate sample size. The study described above delimited sampling in two ways that illustrate this interplay.

First, the study involved a predefined population (prisoners in long-term incarceration) and focused on the general psychological effects of imprisonment. Given that the research question was broad, it was suitable to define a sample by including any individual at the site who had experienced long-term incarceration for at least 10 years. Had the investigation focused on a narrower question, a more differentiated (purposive) sampling strategy would have been required to discern subtle variations.

Second, the study used three standard open-response questions that were posed to each subject in the same way from the beginning to the end of the study and used a *time limitation* on the response of each subject. In this sense, the design was similar to transversal epidemiological studies of prevalence, using a speech sample in the same way that one would use a rating scale across a study population (Bernard, 1994). Limiting the duration of the subject’s response was a critical feature and provided some standardization of the data collected by producing relatively uniform segments. This allowed for the use of computer-assisted linguistic analysis with the ALCESTE program by generating segments of approximately the same length, with roughly equal weighting of each subject’s data.

These potential advantages of a broad sample and a uniform data collection procedure, however, produced tradeoffs regarding theory development. In many qualitative studies, theoretical saturation is defined as the point where the researcher stops collecting new data because the major categories of investigation have been identified across variations in the data obtained up to that point (Bowen, 2008; Glaser and Strauss, 1967). In contrast, in the above study (as in many mixed-methods studies), data collection was completed before beginning data analysis, and saturation was therefore defined as the point where consecutive review of further cases in the sample did not generate new categories.

This does not inherently create problems if the theoretical themes of interest are readily evident in the particular sample. The disadvantage of this procedure, however, is that preset limits on the amount of data collected from each subject may miss theoretically informative material because of the delimited nature of the response conditions. The length of the response time allowed also affects the point at which theoretical saturation is reached. In an ethnographic interview with no fixed time constraint, each subject generally evokes a range of themes and categories representative of the underlying structure of the subjective experience being investigated—in this case, imprisonment. In our study, a 6-minute limit announced in advance likely constrained the responses, leading subjects either to inhibit some statements or to feel time pressure to state what was most important to them quickly. The use of speech segments of predefined length thus provides uniformity of observation, but in a theoretically delimited space.

This interdependency of data format and its eventual theoretical relevance is an example of a larger issue that should be considered before study design, that is, the amount of data that is likely sufficient to answer a question (Robins et al., 2008). In some ways, this consideration is analogous to estimating the required sample size in quantitative studies. Adequate sample size in quantitative terms is determined by the power obtained in using particular measures to examine phenomena with expected effect sizes or prevalences and likely distributions in the group of individuals examined.

A possible guideline by analogy for estimation of sample size in a study focused on the analysis of meaning in transcribed speech samples could thus be approximated as a function of the following:

- the extent to which primary subjects (*e.g.*, patients) are likely to differ widely on the question of interest, a notion analogous to variance;
- the number of other associated groups that will be included in the study (*e.g.*, family members or health care providers), as in a stratified sample;
- the amount of data generated per subject (*e.g.*, length of the interview, total number of interviews if the design is longitudinal), which, for reasons of feasibility, should be inversely proportional to the *number* of subjects in the sample;
- the degree of uncertainty in approaching the topic (how much is already known); and
- something like effect size, which may depend on the degree of subtlety and differentiation desired.

Although it would be intriguing to imagine a formula (analogous to a power calculation) that estimates the needed sample size for a study involving textual data, any such effort must be greeted with skepticism. Nonetheless, these factors should be considered before study design to make a convincing case about why a specific sample size is needed for collection of textual data. This can be performed on a logical basis with explanation of the reasoning, even if estimates are necessarily less structured than in typical quantitative studies.

Reliability and Validity

The notions of reliability and validity have a justifiably hallowed role as indicators of sound practice in quantitative research. Adequate reliability is a prerequisite for validity, because measurement error introduces “noise” into statistical tests that can mask or distort a potential effect within the data (Anastasi, 1934; Cohen, 1990; Cronbach, 1943). Reliability of measurement ensures that the construct is precise, and validity then ensures that the construct is relevant to the nomological net of findings on that topic (Campbell and Fiske, 1959).

Unlike studies using strictly quantitative methodologies, studies involving textual data often do not have readily accessible strategies for formal testing of results. Instead, they have typically been concerned principally with characterizing the structure of meaning and generating hypotheses through a hermeneutic process that takes account of outliers by updating the theory in light of new data, confirming or disaffirming patterns as data collection proceeds. In the case of brief textual data, the issues of reliability and validity can, however, be addressed with a refinement of how these two principles are defined.

Traditional qualitative methods use a distinct set of terms to describe what quantitative researchers call “reliability” and “validity.” Mixed-methods researchers tend to use yet another vocabulary (see Dellinger and Leech, 2007, Table 2). And yet, these two notions are applicable across methodologies, precisely because they are general principles of scientific method: to ensure that knowledge is both relevant to the object of study (validity) and reproducible (reliability). These terms can be applied to textual data in a way that is analogous to how they are applied in positivist approaches.

Although the two principles were both addressed by a single data analysis procedure in the referenced study, we consider reliability and validity separately here to outline how each principle is operative in the approach that was chosen. The data analysis involved a) qualitative content review, and b) use of computer-assisted linguistic software as a distinct mode of alternative analysis of the data. This alternative analysis can be viewed in relation to either reliability or validity. Juxtaposition of two linguistic approaches (content analysis and computer-assisted co-occurrence analysis) likely promoted increased reliability as well as validity of the findings. The two methods each served as a counterpoint to the other, not only confirming the findings but also revealing salient features of the data that might have been missed if only one method had been used. In other words, computer-assisted linguistic analysis is a useful tool precisely because, unlike indexing software, it performs a procedure that is in part automated and thereby provides an alternative method, with the possibility of discovering new themes, categories, and associations not detected by the human reader. Although some authors have claimed that linguistic analysis software can be used as a standalone and that it results in a finding that is more “objective” because of its automated nature (Schonhardt-Bailey, 2005), the approach in the referenced study used the computerized result as an adjunct to guide refinement of the qualitative content review.

Reliability

As Glaser and Strauss (1967, pp 101–102) described in the rationale for the method of constant comparison, formal qualitative interpretive procedures range from immediate quantitative processing of linguistic data to open hermeneutic processes. Some studies using textual data build in reliability in a straightforward manner. Textual data can be coded by multiple analysts for the presence or absence of certain themes, and rules can be developed and tested for consistent application across textual segments, using such indicators as a kappa statistic for both the identification of a segment and the classification of that segment by raters (see Lidz et al., 2000). The resulting data can then be analyzed quantitatively for the particular study, although generalizability of the coding scheme to other studies is not guaranteed.

A more challenging situation arises when textual data are used to generate new theoretical categories or to identify processes that occur in a specific social context, such as in a therapy session or in an individual’s narrative about prison. In these instances, iterative reading of the text, rather than immediate coding and quantification, is valuable for its generative qualities. The principle of reliability is not obvious in such circumstances (Armstrong et al., 1997) and tends to overlap with the assessment of validity. Techniques for measuring interrater reliability in traditional qualitative reviews are generally reserved only for

evaluating the final coding after a coding system has been established. When there are multiple readers performing analyses according to the same procedures in parallel and independently, an equilibrium has to be maintained between a) separation of each reader's results in the interest of later assessing their degree of convergence (reliability) at the end of the analysis and b) the synthesis of interpretations along the way in the interest of a more nuanced hermeneutic process (validity). In the latter circumstance, reproducibility is not at the forefront, and priority is given to discovering a relevant synthesis through collaborative interpretation.

One approach for creating a middle ground between reliability and validity aims is for several readers, specialized in different areas relevant to the research question, to read the text independently, meeting to document points of convergence and divergence through discussion. For example, if a Marxist criminologist, a psychoanalyst, and a biological psychiatrist all identified one or several categories in common, it would be fair to conclude that the categories in common were very strongly present in the data. Where their readings diverged, the burden would be upon the investigators to account for their differences, for example, whether differences arose from distinct training and professional experience or personal views. Bringing multiple interpretations into a convergent account that evolves over multiple readings through open debate about differences can provide one version of interrater reliability that is analogous to procedures used in psychometric ratings.

Validity

Validity in qualitative research is usually formulated in terms of plausibility, credibility, and relevance (Altheide and Johnson, 1998) and involves techniques of triangulation (Glaser and Strauss, 1967). This refers to the use of an alternate means of verifying the result obtained from qualitative content review of the primary data. It can be carried out by collecting data from an alternate sample or using an alternative analysis (Hatch, 2002). As indicated above, this procedure can often be understood as a method of establishing some concordance of interpretation, thus resembling the idea of reliability. Or, if the alternate method of analysis provides a known benchmark for qualitative interpretation, triangulation can be seen as serving the goal of validity. Maxwell (2005) is careful, however, to point out that validity is not a matter of method but rather of the researcher's attentiveness to the presence of specific threats to validity, such as researcher bias or observer effects (reactivity), which modify the object of study as it is undergoing scrutiny. Ultimately, the principles of reliability and validity serve only as guidelines by which the researcher maintains vigilance in the task of establishing a sense of veracity. She or he does so in part through transparency regarding ontological and epistemological assumptions (Altheide and Johnson, 1998; Hatch, 2002).

In the study described, triangulation was approached in two ways: a) through consultation of a second reader and b) through the use of computer-assisted linguistic analysis, which provided information leading to an additional category in the data, as confirmed upon further qualitative review. An additional approach for addressing validity, particularly well-suited to embedded mixed-methods designs where categorical and qualitative data are collected simultaneously on the same subjects (Creswell and Plano Clark, 2007), is to examine outliers to assess the nature of observed disparities in content (Hatch, 2002). Targeted analyses, either quantitative or qualitative, can be performed to demonstrate expected relations among variables or to uncover the potential reasons why some cases do not follow the expected set of relations (see Lidz et al., 1989). This is the approach that was used to identify two subgroups in the study described above and to discern differences in the speech of mentally ill and non-ill inmates.

In summary, reliability and validity can be operationalized in ways that are relevant when data are in the form of words, and the two notions are overlapping principles in this context.

The use of textual data requires particular attention and specificity in the design, data collection, and analysis phases to demonstrate that the principles implied by these two concepts are systematically addressed in a language-based study and that they have resulted in a convincing interpretation of the data collected.

LANGUAGE AS DATA IN PSYCHIATRIC RESEARCH: WHY BOTHER?

Words are fundamentally different from numbers. Examining meaning through textual data provides access to aspects of subjective experience that inherently escape psychometric and neurobiological measurement (see Cusack et al., 2007). Findings from studies using linguistic data, however, should not be taken in isolation but rather juxtaposed, contrasted, and integrated with existing research traditions within psychiatry. We have proposed that, using quantitative principles (*e.g.*, adequate sample size, reliability, validity) as a guide by analogy, researchers can integrate these two traditional approaches to address certain research questions more comprehensively.

The Use of Linguistic Data in Medicine

Medicine's reliance on numerical and categorical data ("quantitative" data analyzed using statistics) is at the heart of its identity as a scientific discipline, but this reliance is a relatively recent development in its longer history. It was not until after the end of World War II, under a particular set of scientific, social, and political conditions, that procedures became established for the evaluation of pharmaceutical treatments through randomized clinical trials (RCTs; Marks, 1997), the current gold standard of proof regarding therapeutic efficacy. The historical reliance on clinical trials and associated statistical methods (*e.g.*, in epidemiology) has led to a strong emphasis on the quantification of results and the use of numerical significance levels as a standard of proof. In this context, the very idea of collecting and analyzing linguistic data would appear to be fraught with imprecision.

Linguistic data regarding subjective experience have been considered relevant to psychosocial processes associated with a disease, such as in studies of quality of life, treatment tolerability and adherence, and patient satisfaction. These methods have made significant inroads in health care research, providing valuable data on subjective aspects of health behavior. Content analysis and mixed methods have long been used by investigators in fields such as nursing, social work, and allied health sciences (O'Cathain, 2009) and more recently in health systems policy (Mays et al., 2005). Some investigators in primary care (*e.g.*, see Eborall et al., 2007) have used brief qualitative sampling methods to study patient preferences and health behavior. A disciplinary difference, however, has fostered the impression that textual methods address issues related to the psychosocial context rather than in the study of disease.

It is not so clear that this division of methods serves psychiatry well (Crawford et al., 2003). In psychiatry, the assessment of subjective experience has the potential to contribute not only to research on health services delivery and quality of life but also to our further understanding of disease processes, treatment, and outcomes. Psychiatric symptoms are embedded in thought processes that are experienced and conveyed most spontaneously through the medium of language. The symptoms and functional problems associated with mental disorders are defined at least partially in terms of how they interact with an individual's interpretation of his world (Cusack et al., 2007; Robins et al., 2005). The very definition of a psychiatric disorder includes self-perception and subjective attribution of meaning, with concomitant influence upon behavior. However, to bring this type of subjectivity into a relevant relationship with existing research in psychiatry, textual data must be analyzed in accordance with principles that are recognizable in current research paradigms.

The Problem of Case-Specific Features

As described above, a robust qualitative research tradition, although receiving less attention, has developed alongside quantitative approaches. Further development of ethnographic qualitative research in psychiatry would surely be valuable in its own right. An equally valuable effort, however, would be to explore the integration and interweaving of qualitative approaches and the methods of clinical trials and epidemiological studies. We have posited that this agenda can be fostered through the inclusion of textual data regarding subjective experience as a way to enrich inquiry, by framing questions in more personalized terms.

In other words, data usually described as “qualitative” may even be effectively integrated into studies whose aims must ultimately be resolved through quantitative assessment. For instance, how might subjective experience be assessed, using speech samples, in studies that require a probability estimate for the prediction of future behavior such as suicide attempts, violence toward others, self-injury, or drug abuse relapse? The study of these behavioral outcomes is fraught with complexities that have led to the development of sophisticated statistical strategies. The collateral cost has been that in order for such studies to be feasible, researchers have tended to rely upon objectively verifiable risk markers such as housing status, age, socioeconomic status, impulsivity, and other variables that are presumed to apply across cases in a homogeneous manner. We would posit that measurement in these research domains could be substantially enriched by identifying and examining subjective features of interest (*e.g.*, how individuals construe threats or attribute blame for life outcomes, how they experience distress). To do so effectively, procedures are needed for the collection and analysis of data that more accurately reflect subjective state and personal meaning, in the form of words. By more precisely elaborating our understanding of the epistemological and methodological differences between the components of a mixed-methods approach, we may identify strategies for combined techniques that are suitable for addressing person-centered research questions in studies that nonetheless require statistical analysis.

A Need for Analogous Procedures

Researchers in medicine may view language-based methods as notably inchoate because there seem to be few solid, transferable rules for judging the soundness of such approaches. The RCT, for instance, has stringent standards for validity, reliability, and significance, along with clear formulae for the calculation of sample size on the basis of significance levels and effect size. Although not appropriate to all research questions, these features of the RCT have made it and its associated methods a beacon for rigor in medical research.

It is certainly not our purpose to minimize the value of these methods. Rather, efforts that build upon this tradition in statistics and translate criteria that are analogous, but epistemologically and heuristically different, for judging the adequacy of language-based methods are needed. Such efforts would refine the ways that textual data can be deployed in psychiatric research and illustrate the value, feasibility, and potential contribution of these approaches. Solid guidelines are needed for judging the soundness of language-based methods, according to reasoning and terminology that are familiar to us as biomedical researchers.

Because of differences in underlying assumptions and preferences, not all researchers will endorse the relevance of this strategy, no matter how carefully it is formulated. At one extreme, quantitative and qualitative approaches can be viewed as mere differences in style that nonetheless share a common logic of inference (King et al., 1994). In contrast, advocates of an incompatibilist position argue that textual or qualitative data and analytic procedures result in information that is so fundamentally different in both ontological and

epistemological terms that they cannot be combined with quantitative findings (Morgan, 2007). In this context, some researchers have argued that mixed methods, by combining traditional qualitative and quantitative approaches, constitute a separate, third methodological domain (Denscombe, 2008), with its own rules, involving a pragmatic integration of the criteria used to assess the soundness of both numerical and textual data analysis techniques (Morgan, 2007). The choice to adopt a mixed-methods approach as outlined above is likely to depend upon its suitability to the research question that the investigator seeks to answer.

The study of subjective meanings through textual data does not intrinsically render it alien to scientific procedures. Although this type of data can and has been analyzed productively using a variety of ethnographic and case study techniques, it is also accessible to scientific principles that are analogous and translatable to those used in quantitative approaches. By focusing on the linguistic nature of data that are often described as “qualitative,” we emphasize that the medium in which the data are collected is susceptible to a distinct type of nonnumerical analysis, but one that can be understood in scientific terms.

If one accepts this premise, there are several challenges ahead. It should be clear that integrating language analysis within psychiatry will require further interdisciplinary collaboration. Although we explored semantic analysis and word cooccurrence in the study referenced above, these modalities require further refinement and expansion. In addition, it would be useful to examine the application of existing methods for syntactic and discourse analyses that identify processes where meaning is distorted or lost. Finally, further clarification is needed regarding intraindividual modes of analysis as they relate to interindividual analyses.

In conclusion, speech in the form of textual data can provide an alternative view of subjective processes that escape measurement through categorical or quantitative variables, or through biological parameters. To better study and integrate meaning in subjective experience with existing approaches in psychiatry, we will need to further advance our ability to logically articulate why the evidence is convincing along recognizable dimensions of scientific argument, in a manner that is analogous, to but distinct from, criteria and arguments in quantitative procedures. Such initiatives would likely enrich one of the major aims of psychiatric research—characterizing the subjective experience of patients and the vagaries of meaning in their lives.

Acknowledgments

The study discussed in this article was supported by grant 2001-148 from the Direction Générale de la Santé (Ministry of Health) and the Direction de l'Administration Pénitentiaire (Ministry of Justice), France. The first author was supported by a grant from the US National Institute of Mental Health (T32 MH16804) and by a VA Advanced Physician Fellowship (MIRECC) during the preparation of the manuscript. Contents do not represent the views of the Department of Veterans Affairs or the US Government.

References

- Altheide, DL.; Johnson, JM. Criteria for assessing interpretive validity in qualitative research. In: Denzin, NK.; Lincoln, YS., editors. *Collecting and interpreting qualitative materials*. Thousand Oaks, CA: Sage Publications; 1998. p. 283-312.
- Alvesson, M.; Skoldberg, K. *Reflexive methodology: New vistas for qualitative research*. 2. Los Angeles, CA: Sage; 2009.
- Anastasi A. The influence of practice upon test reliability. *J Educ Psychol*. 1934; XXV:321–335.
- Andreasen NG, Pfohl B. Linguistic analysis of speech in affective disorders. *Arch Gen Psychiatry*. 1976; 33:1361–1367. [PubMed: 985047]

- Armstrong D, Gosling A, Weinman J, Marteau T. The place of inter-rater reliability in qualitative research: An empirical study. *Sociology*. 1997; 31:597–606.
- Barrett, RJ. *Studies in social and community psychiatry*. Cambridge, England: Cambridge University Press; 1996. The psychiatric team and the social definition of schizophrenia: An anthropological study of person and illness.
- Bernard, HR. *Research methods in anthropology*. 2. Thousand Oaks, CA: Sage Publications; 1994.
- Bowen GA. Naturalistic inquiry and the saturation concept: A research note. *Qual Res*. 2008; 8:137–152.
- Bryman, A. The debate about quantitative and qualitative research: A question of method or epistemology?. In: Boudon, R.; Cherkaoui, M.; Demeulenaere, P., editors. *The European tradition in qualitative research*. London; Thousand Oaks, CA: Sage; 1984/2003. p. 101-106.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull*. 1959; 56:81–105. [PubMed: 13634291]
- Castel, R. *U.S. Medicine and society*. Berkeley, CA: University of California Press; 1988. The regulation of madness: The origins of incarceration in France.
- Chevrie-Muller C, Sevestre P, Segui N. Speech and psychopathology. *Lang Speech*. 1985; 28(pt 1): 57–79. [PubMed: 3897749]
- Cohen J. Things I have learned (so far). *Am Psychol*. 1990; 45:1304–1312.
- Corbin, J.; Strauss, A. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. 3. Thousand Oaks, CA: Sage Publications, Inc; 2008.
- Covington MA, He C, Brown C, Naci L, McClain JT, Fjordbak BS, Semple J, Brown J. Schizophrenia and the structure of language: The linguist's view. *Schizophr Res*. 2005; 77:85–98. [PubMed: 16005388]
- Crawford MJ, Ghosh P, Keen R. Use of qualitative research methods in general medicine and psychiatry: Publication trends in medical journals 1990–2000. *Int J Soc Psychiatry*. 2003; 49:308–311. [PubMed: 14727696]
- Creswell, JW. *Qualitative inquiry and research design: Choosing among five approaches*. 2. Thousand Oaks, CA: Sage Publications; 2007.
- Creswell, JW.; Plano Clark, VL. *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage; 2007.
- Cronbach LJ. On estimates of test reliability. *J Educ Psychol*. 1943; 34:485–494.
- Cusack KJ, Grubaugh AL, Yim E, Knapp RG, Robins CS, Frueh BC. Are there racial differences in the experience of harmful or traumatic events within psychiatric settings? *Psychiatr Q*. 2007; 78:101–115. [PubMed: 17345158]
- Dellinger AB, Leech NL. Towards a unified validation framework in mixed methods research. *J Mixed Methods Res*. 2007; 1:309–332.
- Denscombe M. Communities of practice: A research paradigm for the mixed methods approach. *J Mixed Methods Res*. 2008; 2:270–283.
- Donmoyer, R. Generalizability and the single-case study. In: Gomm, R.; Hammersley, M.; Foster, P., editors. *Case study method: Key issues, key texts*. London; Thousand Oaks, CA: Sage; 1990/2000. p. 45-68.
- Eborall H, Davies R, Kinmonth AL, Griffin S, Lawton J. Patients' experiences of screening for type 2 diabetes: Prospective qualitative study embedded in the addition (Cambridge) randomised controlled trial. *BMJ*. 2007; 335:490. [PubMed: 17762000]
- Ehrenreich, B.; English, D. *For her own good: 150 years of the experts' advice to women*. 1. Garden City, NY: Anchor Press; 1978.
- Estroff, SE. *Making it crazy: An ethnography of psychiatric clients in an American community*. Berkeley, CA: University of California Press; 1981.
- Falissard, B. *Mesurer la subjectivité en santé: Perspective méthodologique et statistique [Measuring subjectivity in health: Methodological and statistical perspectives]*. Paris: Masson; 2001.
- Foucault, M. *History of madness*. London: Routledge; 1961/2006.
- Gauchet, M.; Swain, G. *Madness and democracy: The modern psychiatric universe. New french thought*. Princeton, NJ: Princeton University Press; 1980/1999.

- Glaser, BG.; Strauss, AL. The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine Publishing; 1967.
- Goldstein, J. Console and classify: The French psychiatric profession in the nineteenth century. 2. Cambridge, England: Cambridge University Press; 2002.
- Gomm, R.; Hammersley, M.; Foster, P. Case study and generalization. In: Gomm, R.; Hammersley, M.; Foster, P., editors. Case study method: Key issues, key texts. London; Thousand Oaks, CA: Sage; 2000. p. 98-115.
- Gottschalk LA, Stein MK, Shapiro DH. The application of computerized content analysis of speech to the diagnostic process in a psychiatric outpatient clinic. *J Clin Psychol*. 1997; 53:427–441. [PubMed: 9257220]
- Hatch, JA. Doing qualitative research in education settings. Albany, NY: State University of New York Press; 2002.
- Haynes SN, Mumma GH, Pinson C. Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clin Psychol Rev*. 2009; 29:179–191. [PubMed: 19217703]
- Hoffman RE, Stopek S, Andreasen NC. A comparative study of manic vs schizophrenic speech disorganization. *Arch Gen Psychiatry*. 1986; 43:831–838. [PubMed: 3753163]
- Hohmann AA, Shear MK. Community-based intervention research: Coping with the “noise” of real life in study design. *Am J Psychiatry*. 2002; 159:201–207. [PubMed: 11823259]
- Horwitz, AV. Creating mental illness. Chicago: University of Chicago Press; 2002.
- Horwitz, AV.; Wakefield, JC. The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder. Oxford: Oxford University Press; 2007.
- Image. Alceste: Un logiciel d'analyse de données textuelles. 4.9. Toulouse, France: Image; 2007.
- Jenkins JH, Carpenter-Song EA. Awareness of stigma among persons with schizophrenia: Marking the contexts of lived experience. *J Nerv Ment Dis*. 2009; 197:520–529. [PubMed: 19597360]
- Jodelet, D.; Duveen, G. Madness and social representations: Living with the mad in one french community. Medicine and society. Berkeley, CA: University of California Press; 1991.
- Johnson RB, Onwuegbuzie AJ, Turner LA. Towards a definition of mixed methods research. *J Mixed Methods Res*. 2007; 1:112–133.
- Kendler KS. Toward a philosophical structure for psychiatry. *Am J Psychiatry*. 2005; 162:433–440. [PubMed: 15741457]
- King, G.; Keohane, RO.; Verba, S. Designing social inquiry: Scientific inference in qualitative research. Princeton, NJ: Princeton University Press; 1994.
- Krauss S. A tripartite model of idiographic research: Progressing past the concept of idiographic research as a singular entity. *Soc Behav Pers*. 2008; 36:1123–1140.
- Kuperberg GR, Kreher DA, Ditman T. What can event-related potentials tell us about language, and perhaps even thought, in schizophrenia? *Int J Psychophysiol*. 2010; 75:66–76. [PubMed: 19765622]
- Lane, C. Shyness: How normal behavior became a sickness. New Haven, CT: Yale University Press; 2007.
- Lidz, CW.; Meisel, A.; Zerubavel, E.; Carter, M.; Sestak, RM.; Roth, LH. Informed consent: A study of decision making in psychiatry. New York: Guilford Press; 1984.
- Lidz CW, Mulvey EP, Appelbaum PS, Cleveland S. Commitment: The consistency of clinicians and the use of legal standards. *Am J Psychiatry*. 1989; 146:176–181. [PubMed: 2912259]
- Lidz CW, Mulvey EP, Hoge SK, Kirsch BL, Monahan J, Bennett NS, Eisenberg M, Gardner W, Roth LH. Sources of coercive behaviours in psychiatric admissions. *Acta Psychiatr Scand*. 2000; 101:73–79. [PubMed: 10674953]
- Lim L, Nathan P, O'Brien-Malone A, Williams S. A qualitative approach to identifying psychosocial issues faced by bipolar patients. *J Nerv Ment Dis*. 2004; 192:810–817. [PubMed: 15583501]
- Luhrmann, TM. Of two minds: The growing disorder in American psychiatry. 1. New York: Knopf; 2000.
- Lunbeck, E. The psychiatric persuasion: Knowledge, gender, and power in modern America. Princeton, NJ: Princeton University Press; 1994.

- Marks, HM. The progress of experiment: Science and therapeutic reform in the United States. New York: Cambridge University Press; 1997. p. 1900-1990.
- Maxwell, JA. Qualitative research design: An interactive approach. Thousand Oaks, CA: Sage Publications; 2005.
- Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. *BMJ*. 2000; 320:50–52. [PubMed: 10617534]
- Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *J Health Serv Res Policy*. 2005; 1(Suppl 10):6–20. [PubMed: 16053580]
- Morgan DL. Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *J Mixed Methods Res*. 2007; 1:48–76.
- Newman S, Mather VG. Analysis of spoken language of patients with affective disorders. *Am J Psychiatry*. 1938; 94:913–942.
- O’Cathain A. Mixed methods research in the health sciences: A quiet revolution. *J Mixed Methods Res*. 2009; 3:3–6.
- Onwuegbuzie AJ, Leech NL. A call for qualitative power analyses. *Qual Quantity*. 2007; 41:105–121.
- Oxman TE, Rosenberg SD, Schnurr PP, Tucker GJ. Diagnostic classification through content analysis of patients’ speech. *Am J Psychiatry*. 1988; 145:464–468. [PubMed: 3348449]
- Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: Our words, our selves. *Annu Rev Psychol*. 2003; 54:547–577. [PubMed: 12185209]
- Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *BMJ*. 2000; 320:114–116. [PubMed: 10625273]
- Robins CS, Sauvageot JA, Cusack KJ, Suffoletta-Maierle S, Frueh BC. Consumers’ perceptions of negative experiences and “sanctuary harm” in psychiatric settings. *Psychiatr Serv*. 2005; 56:1134–1138. [PubMed: 16148329]
- Robins CS, Ware NC, dos Reis S, Willging CE, Chung JY, Lewis-Fernandez R. Dialogues on mixed-methods and mental health services research: Anticipating challenges, building solutions. *Psychiatr Serv*. 2008; 59:727–731. [PubMed: 18586988]
- Roe D, Hasson-Ohayon I, Kravetz S, Yanos PT, Lysaker PH. Call it a monster for lack of anything else: Narrative insight in psychosis. *J Nerv Ment Dis*. 2008; 196:859–865. [PubMed: 19077852]
- Sajatovic M, Jenkins JH, Cassidy KA, Muzina DJ. Medication treatment perceptions, concerns and expectations among depressed individuals with type I bipolar disorder. *J Affect Disord*. 2009; 115:360–366. [PubMed: 18996600]
- Schofield, JW. Increasing the generalizability of qualitative research. In: Gomm, R.; Hammersley, M.; Foster, P., editors. *Case study method: Key issues, key texts*. London; Thousand Oaks, CA: Sage; 1990/2000. p. 69-97.
- Schonhardt-Bailey C. Measuring ideas more effectively: An analysis of Bush and Kerry’s national security speeches. *Political Sci Politics*. 2005:701–711.
- Showalter, E. *The female malady: Women, madness, and English culture, 1830–1980*. 1. New York: Pantheon Books; 1985.
- Stamm TA, Cieza A, Coenen M, Machold KP, Nell VP, Smolen JS, Stucki G. Validating the International Classification of Functioning, Disability and Health comprehensive core set for rheumatoid arthritis from the patient perspective: A qualitative study. *Arthritis Rheum*. 2005; 53:431–439. [PubMed: 15934102]
- Teddlie C, Yu F. Mixed methods sampling: A typology with examples. *J Mixed Methods Res*. 2007; 1:77–100.
- Toch, H. *Men in crisis: Human breakdowns in prison*. Chicago: Aldine Publishers; 1975.
- Toch, H. *Living in prison*. New York: Free Press; 1996.
- Toch, H.; Adams, K.; Grant, JD. *Acting out: Maladaptive behavior in confinement*. 1. Washington, DC: American Psychological Association; 2002.
- Weintraub W, Aronson H. The application of verbal behavior analysis to the study of psychological defense mechanisms: Methodology and preliminary report. *J Nerv Ment Dis*. 1962; 134:169–181. [PubMed: 14005787]

- Widlöcher D. La méthode du cas unique. Monographies Psychopathologie. 1999; 1:191–200.
- Yang S, Kadouri A, Revah-Levy A, Mulvey EP, Falissard B. Doing time: A qualitative study of long-term incarceration and the impact of mental illness. *Int J Law Psychiatry*. 2009; 32:294–303. [PubMed: 19619895]