

Mergenthaler, E., & Kächele, H. (1992). Text Archive Facilities and their use in Psychotherapy Research. In N. Miller L. Luborsky J. Barber & J. Docherty (Ed.), *Handbook of dynamic psychotherapy research and practice* (pp. New York: Basic Books.

## **Text Archive Facilities and their use in Psychotherapy Research**

Erhard Mergenthaler & Horst Kächele

In the first edition of the *Handbook of Psychotherapy and Behavior Change* Luborsky and Spence (1971) already were asking for empirical studies based on primary data. Primarily these are audio- and video-recordings from therapeutic sessions and transcriptions of these. Meanwhile all over the places where psychotherapy research takes place this desideratum became reality. Bunches of tapes and transcripts have been produced. Lots of them have been used only once. The data sets are no longer available or accessible.

A significant step was taken when a set of transcripts of psychoanalytic sessions was made available by Hartvig Dahl to a broader group of scientists. Various ways of empirical assessment could be applied to the same session (e.g., Dahl, Kächele, and Thomä, 1988). Dahl also developed one of the first guides for transcription (Dahl, 1979) which was utilized by scholars thereafter. In Europe, especially in the German speaking countries Kächele and Thomä shortly after work started out to audio-tape psychoanalytic and psychotherapeutic sessions and then having them shared with others for basic research. This did not remain restricted to the field of psychotherapy, but linguists also became interested in this type of textual material. The transcription standards, developed for German language by Mergenthaler, came in widespread use. As a result of these interdisciplinary endeavors, these guides to transcription were made to fulfill the most likely needs of a researcher in psychotherapy as well as those of a linguist. Additionally these rules were made for full compatibility with computer use. Recently Mergenthaler and Stinson (1992) presented the English version of these transcription standards. These already are used in various places.

The idea of an archive centralizing data sources of this kind was born in 1979 by Kächele and Mergenthaler. With the funding of the German Research Foundation this idea could become reality within the following nine years. In 1989 Sherwood Waldron initiated the Psychoanalytic Research Consortium as an archive for psychoanalytic recordings and transcriptions from North America.

In the following chapters we will give a brief overview of existing facilities today. The emphasis however we will rather put on the regulations of how the archive is administered than on an exact listing of what is available at the moment. The numbers quickly change and easily can be updated by a call or written inquiry to the respective place. We will take somewhat more space for the lending policy, the means of ensuring privacy and authorship. For those who are interested in more detail we have the addresses at the end of this contribution. Each of the mentioned archive or persons will give more information.

## **2. Major archives**

Up to date three major archives are known. Two of them, the Psychoanalytic Research Consortium (PRC) and the Center for the Study of Neuroses (CSN) collect English material only. The Ulm Textbank (UTB) takes care of both, English and German recordings and transcriptions. More details will be given in the following presentations of each of them. However we do not present a list or catalog of all holdings here due to updates which would be necessary in short periods of time. Rather we highlight some of the materials available with each of the archives.

### **2.1 The Ulm Textbank**

The Ulm Textbank (UTB) started in the late seventies and has today worldwide the largest collection of reports, test protocols, tapes and transcripts. It is part of the University of Ulm, Germany and is run by the Section on Computer Science in Psychotherapy (Director: E. Mergenthaler) and affiliated to the department of Psychotherapy (Clinical director: H. Kächele). The UTB offers a full spectrum of services: 1) Intake of relevant material, 2) lending of transcripts, tapes or data files on diskettes, 3) Computer-aided analyses of textual data, and 4) counseling for the planning of empirical studies. Besides that at UTB methodologies are developed to economically administer an archive as such (text bank managing system) and to automatically analyze textual data. Based on the huge amount of text in the archive characteristics of various linguistic phenomena also are determined and made public.

The holdings of the Ulm Textbank comprise material of English and German origin. The German language data base now includes extensive samples from four psychoanalytic cases. Individual sessions from nine other psychoanalytic therapies are also included. The database of initial interviews includes about 500 intake interviews and is referenced according to the sex of the patient or therapist and whether the diagnosis is neurosis or psychosomatic disturbance. This body of texts steadily is being enlarged with special attention to the patient variables of sex, diagnosis, social class, and age, and the therapist variables of experience and kind of psychotherapy. The English portion of texts in the Ulm Textbank is

made up of 40 sessions stemming from the Penn Psychotherapy Project and representing two early and two late sessions each from 10 improved and 10 non-improved patients in psychodynamic therapy. For the following texts requests might be sent via the Ulm Textbank, however they are not regular parts of the archive yet and restrictions may apply. In particular these are 15 stratified random hours each from 15 analytic cases (total of 225 hours). Another 25 sessions belong to an analytic case selected as 10 "work", 10 "resistance", and 5 "middle" hours. Finally there are 185 sessions from another analytic case.

The kind of text included in the Textbank is determined by the goals, questions, and scientific interests of the supporting and other institutions. For the Department of Psychotherapy at Ulm University, this means an emphasis on the establishment of an empirical basis for research in the field of psychotherapy, and to enhance teaching. The latter takes the form of demonstration materials for the education of medical students and the use of verbatim transcripts in the clinical education and supervision of resident physicians and psychologists.

The UTB is open for everyone engaged in transcript or record based research. The intake of material is done in one of three levels the donator can freely choose from: 1) The material is given to the UTB just for internal use as developing characteristics for linguistic variables. In this case the material is not subject to be lent nor to become printed or edited. Thus UTB can guarantee full confidentiality for this kind of material. 2) The material is given to the UTB as available for lending. However, before the material (a copy of) will be given out, the donator will be informed by the UTB and asked for permission. As source of information UTB asks the interested scholar for an explanation of the study the material is wanted for and for a proof of institutional affiliation. 3) The material is given to the UTB at its free disposal. The same conditions for holding the information apply as in 2), however, the administrative amount of work is less.

Whatever level a donator has chosen, she or he will be informed when a study using his material has been completed. If available a copy of the publications as a result of the studies will be provided.

Those who are interested in lending some material will have to submit an overview of what they intend to do with the material. The staff of the UTB will evaluate these applications with special regard to confidentiality and ethics. If acceptable a form will be sent to the respective person which has to be signed and returned before any material is given out. Essentially this declaration restricts the use of the material loaned to the person only and for the outlined research purposes only. There is a regulation that it is not allowed to have copies made from the material loaned and that it has to be returned in full when the research is

done. If a publication is the outcome of the research a reprint of it has to be given to the UTB. Before printing, quotations from the material loaned have to be checked with the staff of the UTB for confidentiality.

One of the favorite services of the UTB is the computer aided analysis of the textual material. The Textbank has a variety of methods which can be applied in a tailor-made way. This also includes offering advice when new studies are planned. Due to the long experience at UTB studies can often be quickly and effectively planned in collaboration with the staff of the UTB. Information about the material is available, such as a catalog of all available material and a list of all studies with reference to the material that has been finished to date.

The services of the UTB are available essentially at very low cost. Charges are made only for the costs of the labor-intensive tasks, such as the transcription of tape recordings and small charges for materials.

## **2.2 The Psychoanalytic Research Consortium**

The Psychoanalytic Research Consortium (PRC) according to its chairman, Sherwood Waldron, is a non-profit corporation since 1989 whose purpose is to collect recordings of psychoanalysis and other psychoanalytically oriented psychotherapies, and to catalogue and safeguard the tapes and other materials relating to these treatments. The PRC also prepares suitable selections of such materials for distribution to qualified psychoanalytic researchers.

The approval of researchers for access to materials basically is comparable to what has already been outlined for the Ulm Textbank. The purpose of the PRC is to support psychoanalytic research especially.

Comparable standards also are applied for the preparation of materials. Transcription is done according to the same standards. This implies use of similar techniques for preserving confidentiality and equivalent restrictions on use and publication of materials.

The PRC has available or has access to recordings of more than 25 analyses by nine different analysts. The text of one analysis of 324 hours is completely available on computer. The other materials are transcribed to varying degrees. A number of psychotherapies conducted by analysts are also available. Because the PRC is an autonomous institution it is dependent upon fees from users for operating funds.

## **2.3 Center for the Study of Neuroses**

In contrast to the Ulm Textbank and the Psychoanalytic Research Consortium, the Center for the Study of Neuroses does not give away textual material but rather wants to encourage scientists at various sites to participate in collaborative research endeavors. Requests must be for a project proposal approved by the Directors and the Coordinating Committee of

the Program for the Study of Conscious and Unconscious Mental Processes at UCSF supported by the John D. and Catherine T. MacArthur Foundation. Work must be done with a designated Liaison Scientist, whose responsibility is to facilitate collaboration and to ensure proper use of materials.

Among other materials like Running Notes, Process Notes, Videotapes, Audiotapes the archive has 3 fully transcribed and time coded brief therapies with patients suffering from pathological grief. Additional materials as experimental data, psychological testing scores, and recordings of research sessions, evaluation sessions, and follow-up sessions including physiological data are available. Furthermore there are transcripts from another 30 grief cases.

#### **2.4. Other collections and archives**

There are a couple of other places with collections of recordings and transcriptions. Both, UTB and PRC will forward inquiries if their material fits the specific needs.

### **3 Methodologies for analyzing transcript data**

Starting from a semiotic view, language is understood as a system of symbols whose structure is determined according to rules based on the relationship between form and content. Accordingly, it is possible to distinguish between formal, grammatical, and substantive measurements. Each of these types of measurements can be further subdivided according to whether it can be applied to a speaker's text or to the entire speech activity in a conversation (the dialogue). It is therefore possible to speak of monadic or dyadic values and to distinguish among these types of measurements according to the kind of data they utilize. Best known are simple frequencies of occurrence, which form the basis for ratios and distributions.

It should also be noted that some of the approaches for formal and grammatical measurements presume substantive knowledge of, for example, the denotative meaning of a word. The contrast with substantive measurements stems from the fact that the required knowledge does not come from the psychoanalytic research field itself, but from the realm of methodologies; that is, from linguistics or information science.

The formal measurements can generally be determined in a simple manner. In computer-aided approaches, only the capacity to segment a sequence of symbols (letters, numbers, and special symbols) into words and punctuation are necessary. Such formal measurements and indications of their applicability include: Text size (Tokens), Vocabulary (Types), Type-Token Ratio, Redundancy.

The simplest and most elementary formal measurement is that of the numbers of words spoken by the analyst and patient. Kächele (1983) found that in a successful psycho-

analytic treatment there was no correlation across 130 sessions in the number of words spoken by analyst and patient. In a rather unsuccessful treatment by the same analyst, these word counts were significantly correlated.

The redundancy of a text is a measure adopted from information theory. Spence (1968) proposed some important ideas about psychodynamic redundancy, without testing these ideas empirically. In addition, he formulated a series of hypotheses about the course that redundancy takes in psychoanalytic treatment. Kächele and Mergenthaler (1984) confirmed one of these hypotheses, that the repetitiousness of a patient's speech increased over the course of treatment. The therapist's redundancy, in contrast, remained constant.

Grammatical measures address such features as interjections, and word class (part of speech). The grammatical measures require the researcher to have linguistic knowledge about the language being studied; for example, the grammar of English. The programming and pre-coding tasks in the computer-aided procedures are considerable. Moreover, many kinds of questions still cannot be correctly processed automatically. An example is lemmatization, which can assign 50% to 95% of all word forms, depending on the kind of text, to the correct lemma. The psychoanalytic interview, a form of speech with the many syntactically deviant forms (such as incomplete words and sentences) that characterize spoken and spontaneous speech, pose unique challenges. Accordingly, there are hardly any computer-aided studies of psychoanalytic texts using grammatical measures.

The dependence of word choice on word type and semantic class was demonstrated by Busemann (1925) in investigations of children's speech. He spoke of an "active" and a "qualitative" style with regard to verbs and adjectives. He showed that these differences in style are only slightly dependent on the subject being discussed and stem from personality variables. Using a computer-aided approach to the text of a psychoanalytic interview, Mergenthaler and Kächele (1985) showed that the realization of a word form within the text may depend on the subject matter. However, this microanalytic view does not exclude the possibility that, viewed at the macro level, personality-dependent variables are effective as described by Busemann (1925).

Substantive measures require, in addition to the knowledge mentioned above, detailed expert knowledge of a theory and its area of application. Computer-aided procedures are only able to provide approximate results and are limited to narrowly defined constructions (Lolas, Mergenthaler, and von Rad 1982). New approaches in information science, especially in the field of artificial intelligence, could achieve a breakthrough by establishing data bases in conjunction with a system of rules. Still, however, practical tools are not available.

In a pioneering content analytic study, Dahl (1972) was able to trace the downhill course of 363 hours during a 2 1/2 year segment of an unsuccessful psychoanalysis and to convincingly categorize 25 sessions as 10 extreme "work" hours, 10 extreme "resistance" hours, and 5 directly in the middle of the range. Then, using single words derived from the Harvard III dictionary categories, he was also able to demonstrate word clusters that manifestly appear to reflect oedipal and other unconscious conflicts (Dahl, 1974). Reynes, Martindale and Dahl (1984) used the Regressive Imagery Dictionary (RID) to compare this same patient's 10 working hours and 10 resistance hours. The working hours were characterized by increases in the dictionary categories that assessed primary process language, and the resistance hours were characterized by increases in the secondary process category scores. This agrees with Freud's earlier attribution of defensive functions to the secondary processes (Bucci, 1988).

Large continuous segments as well as selected sections of treatment transcripts, may thus be examined using computer-aided text analysis as a tool in psychoanalytic process research. Further progress requires that methods be developed more extensively, that basic research be conducted, and that techniques from related scientific disciplines such as information science and linguistics be integrated.

#### **4 Available software**

Computer-aided analysis of psychotherapy transcripts needs software which is capable of distinguishing at least the speaking of the therapist and the patient. Furthermore turn at speaking, hour, treatment may be other desired criteria for defining the units of analysis. Other desired segmentation criteria might be the on- and off-set of dreams, early memories or specific episodes. Text analytic software, which used to be developed in the field of the social sciences or literary research (e.g., General Inquirer, OCP) does not account for such a kind of segmentation. Mostly they support structures like chapter, paragraph, or line. The analyses which could be performed with these systems mostly were restricted to delimited areas like building concordances (OCP) or performing content analysis (General Inquirer).

The development of TAS/C (Mergenthaler, 1987), a text archive and analysis system, came out of the needs of psychotherapy research and represents a useful selection of methods and segmentation techniques. As an input TAS/C accepts transcripts according to Dahl or Mergenthaler/Stinson standards. Tas/C is a system that allows for maintaining a text archive and analyzing the texts stored. It takes care of the dialogue structure of a text and is capable of administering marked segments, such as dreams, memories, episodes. According to the scheme given in the previous chapter TAS/C can be used to do formal, grammatical

and content analytic studies. There are also tools to assist when developing content analytic dictionaries, for disambiguating texts and for printing. Interfaces to statistical and graphical software exist. As a technical environment TAS/C needs a computer running the UNIX operating system. Meanwhile this is available also for personal computers using a 386 or 486 processor.

## 5. Summary

Text archives have become increasingly crucial to the rapid maturation of dynamic psychotherapy research. We have listed the main archives and have given a few examples of their applications, particularly to linguistic analyses. But although computer analyses have often been used for limited linguistic analyses many of these as illustrated, have bearing on important dynamic issues.

## 6 References

- Bucci, W. (1988). Converging evidence for emotional structures: Theory and method. In H. Dahl, H. Kächele & H. Thomä (Ed.), *Psychoanalytic process research strategies* (pp. 29-49). Berlin Heidelberg New York: Springer-Verlag.
- Busemann, A. (1925). *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik*. Jena.
- Dahl, H. (1972). A quantitative study of a psychoanalysis. In R. R. Holt & E. Peterfreund (Ed.), *Psychoanalysis and contemporary science* (pp. 237-257). New York: Macmillan.
- Dahl, H. (1974). The measurement of meaning in psychoanalysis by computer analysis of verbal context. *J Am Psychoanal Assoc*, 22, 37-57.
- Dahl, H. (1979). *Word frequencies of spoken American English*. Essex CN: Verbatim.
- Dahl, H., Kächele, H., & Thomä, H. (Eds.) (1988). *Psychoanalytic process research strategies*. Heidelberg Berlin New York: Springer-Verlag.
- Kächele, H. (1983). Verbal activity level of therapists in initial interviews and long-term psychoanalysis. In W. Minsal & W. Herff (Ed.), *Methodology in psychotherapy research* (pp. 125-129). Frankfurt: Lang.
- Kächele, H., & Mergenthaler, E. (1984). Auf dem Wege zur computerunterstützten Textanalyse in der psychotherapeutischen Prozeßforschung. In U. Baumann (Ed.), *Psychotherapie: Makro-/Mikroperspektive* (pp. 223-239). Göttingen: Verlag für Psychologie Dr. C. J. Hogrefe.
- Lolas, F., Mergenthaler, E., & Rad, M. v. (1982). Content analysis of verbal behaviour in psychotherapy research: A comparison between two methods. *BR J Med Psychol*, 55, 327-333.



Luborsky, L., & Spence, D. P. (1971). Quantitative research on psychoanalytic therapy. In A. E. Bergin & S. L. Garfield (Ed.), *Handbook of psychotherapy and behavior change: An empirical analysis* (pp. 408-438). New York: Wiley.

Mergenthaler, E. (1987). TAS/C User Manual. University of Ulm: Ulm Textbank.

Mergenthaler, E., & Kächele, H. (1985). Changes of latent meaning structures in psychoanalysis. *Sprache und Datenverarbeitung*, 9(2), 21-28.

Mergenthaler, E., & Stinson, C. H. (1992). Psychotherapy Transcription Standards. *Psychotherapy Research*, 2(1), 58-75.

Reynes, R., Martindale, C., & Dahl, H. (1984). Lexical differences between working and resistance sessions in psychoanalysis. *J Clin Psychol*, 40, 733-737.

Spence, D. P. (1968). The processing of meaning in psychotherapy: Some links with psycholinguistics and information theory. *Behav Sci*, 13, 349-361.

#### Addresses

Ulm Textbank, University of Ulm - Klinikum  
Am Hochstrass 8  
W-7900 ULM, Germany

Psychoanalytic Research Consortium, Inc.  
1235 Park Avenue, Suite 1B  
New York, NY 10128