H. Kächele, H.J. Grünzig & E. Mergenthaler

A computer-based assessment of long-term process trends[1]

Introduction

The Ulm study group on psychoanalytic process research works since many years on the investigation of long-term psychoanalytic processes. The methodology we developed for this ambitious undertaking consists of four approaches each of them focusing on different materials and such on different levels of conceptualization:

I clinical case study
II systematic clinical descriptions
III clinical judgment procedures
IV computer-assisted content analysis

The following investigation combines approaches three and four in an effort to demonstrate that the computer content analysis is able to significantly increase our descriptive power in process research.

Material

The material for this study consists of 130 session (out of 630 sessions) that were transcribed and stored in the ULM TEXTBANK. The distribution of the sample over the course of the psychoanalytic treatment is a stratified time sample; all 25 sessions 5 sessions are included in the study (session 1-5, 26-30, 51-55, ..., 626-630). The sample thus represents a fifth of the whole treatment process. The decision not to use a random time sample depends on our knowledge that in between sessions there is considerable variation of topic and content. Averaging over a sample of fives sessions gives a fair clear account of the present thematic focal structure.

The patient Christian Y[2] whose analytic treatment performed by therapist 001 (code of the Um Textbank) has been the object of this study[3], was a young man suffering from a very severe anxiety neurosis with a narcissistic personality structure, had to be

[1] The study is partially based of "Habilitationsschrift" of one of the authors (Kächele 1976).

[2] The patient Christian Y has been clinically described in Thomä & Kächele (1994b).

[3] The time between termination of this part of the treatment and performance of this study was about three years.

treated as inpatient for nearly two years, so debilitating were his states of free floating anxiety and panic attacks which were accompanied by bouts of very extreme tachycardia.

Study 1: guided clinical judgment of transference and anxiety

In the first study three clinically trained judges (the treating senior analyst, a second junior analyst and a clinical psychologist) rated 55 sessions independently for the intensity and degree of following clinically defined concepts:

= positive transference
= negative transference
= castrations anxiety
= guilt anxiety
= shame anxiety
= separation anxiety
= diffuse anxiety

The definition and operational description fo the concept for the two transference scales focused on the prevailing affective tone of the relationship of the patient to the analyst. The definitions of the anxiety scales were taken from Gottschalk and Gleser's (1969) textbook (a) and manual (b).
The Likert scales for the rating of the verbatim transcripts were 0 - 3: a set secondary scales for the degree of consciousness were -3 to +3. The value of both scales were integrated by a formal algorithmic procedure which followed the reasoning that the estimation of degree of intensity is of higher hierarchy and the estimation of degree of consciousness functions as a moderator variable on the degree of intensity.
By the following formula a set of integrated values was achieved:
Formula: 10 x Xintensity + Xconsciousness = Xcombined.
With these integrated values interrater coefficients were computed:

Table 1. Inter-Rater Reliability among Raters A, B & C

	A / B	A / C	B / C	mean r
positive transference	+ .44	+ .59	+ .32	+ .45
negative transference	+ .47	+ .35	+ .42	+ .41
castrations anxiety	+ .31	+ .38	+ .39	+ .33
guilt anxiety	+ .37	+ .34	+ .31	+ .34
shame anxiety	+ .29	+ .04	+ .07	+ .13
separation anxiety	+ .55	+ .53	+ .30	+ .46
diffuse anxiety	+ .59	+ .53	+ .49	+ .54

N = 55 p = .05, if r = .27; p = .01, if r = .35 Pearson r

Testing for homogenity of the correlation coefficients of the three pair of raters led in no case to a deviation.

The reliability of judgment among the three raters achieves statistical level of significance in most of the cases - except the concept ‚shame anxiety' deviates considerably which could be traced to a different understanding of the three raters - whereas the numerical values on agrement are less than satisfying. However, taking into consideration the task of judgment - to read verbatim protocols of many pages - 10 to 20 pages per session - our results are comparable with those of Bellak and Smith (1956) and Strupp, Chassan and Ewing (1966); these authors have pointed out that the reliablity of judging complex clinical concepts strongly depends on the degree of inference.

In face of the numerical low values of rater consensus we decided not to use the arithmetic mean of the three raters for each concept, but performed a factor analysis to extract the common dimensions of all three raters. Out of the five factor solution we present three factors that yielded clinically sensible composites:

Table 2 Three factors out of the rotated factor matrix

Factor 2: separation anxiety with positive transference

rater	concept	loading
A	postive transference	+ .67
A	separation anxiety	+ .79
B	positive transference	+ .43
B	separation anxiety	+ .59
C	positive transference	+ .79
C	separation anxiety	+ .66

Factor 2: diffuse anxiety with negative transference

A	negative transference	+ .67
A	diffuse anxiety	+ .74
B	negative transference	+ .61
B	diffuse anxiety	+ .64
C	negative transference	+ .35
C	diffuse anxiety	+ .63

In this five-factor solution none of the castration anxiety variables show any significant loadings on any of the five factors. We therefore chose out of a nine-factor solution a specific castration factor for further work.

Factor: castration anxiety

A	castration anxiety	+ .70
B	castration anxiety	+ .68
C	castration anxiety	+ .72

Ordering the clinical judgments by the factorial structure brings out three dimensions that are meaningfully to interpret: factor 2 describes the close connection between ‚separation anxiety' and ‚positive transference', factor 3 describes the close connection between ‚diffuse, objectless anxiety' and ‚negative transference', the specific factor ‚castration anxiety' unifies the judgments on the concept into a higher organized scheme. The factors 1 and 4 Looking back to the procedure of judging such a vast amount of clinical material as we did in this step of the investigation we are more than ever of the opinion that this approach is too cumbersome to yield fruitful long-term committment[4]. This led to the development of a second approach in this investigation.

Study 2: Prediction of clinical judgments by computer content analysis

Our interest in computer assisted content analysis (CACC) has been instigated by Dahl's (1972, 1974) and Spence's (1973, 1976)) exciting investigations with this tool.

We first implemented the PI-1 program described by Spence (1969) and found it useful for dealing with small amount of text material (Kächele et al., 1973). In 1975 we got hold of the program EVA („Elektronische VerbalAnalyse" developed by Holzscheck) which had been devised for the analysis of newspaper headlines. The amount to be

[4] Further details see Grünzig, Kächele and Thomä, 1978

processed with psychoanalytic material made a revision necessary which was performed by Mergenthaler in the following years (so called EVA-Ulm).

Question

We asked whether it is possible to predict the judgment of clinical concepts by the linear combination of category scores. Our aim was to find a way to implement computer assisted content analysis as a scanning procedure; thus would be able to run through large data bases and find sessions with particular clinical constellations.

Method

As analyzing tool for this investigation we used an adaption of the Harvard III Psychosociological Dictionary (Stone, 1966); for technical reasons we restricted the entries in the various categories of the dictionary to the nouns that we identified by a vocabulary analysis of both patient and analyst. The reliability of the dictionary we devised coded 70% of all occurring nouns in the verbatim protocols with a minor difference between the degree of text-coverage of the patient and the analyst's productions. The degree of similarity of the categories between patient and analyst over the course of treatments is considerably; the following table 4 shows the correlation of some selected categories over the whole treatment (N = 130 sessions):

Table 4 Correlation of selected dictionary categories between patient and analyst over the course of treatment

variable	category name	correlation coefficient
1	MALE ROLE	+ .64
2	FEMALE ROLE	+ .59
3	NEUTER ROLE	+ .39
4	JOB ROLE	+ .39
6	BODY PART	+ .48
9	TOOL	+ .20
10	NATURAL OBJECT	+ .45
11	NON SPECIFIC OBJECT	+ .06
12	TIME REFERENCE	+ .34
15	SOCIAL PLACE	+ .28
16	NATURAL WORLD	+ .23
17	IDEAL WORLD	+ .38
18	DEVIATION	+ .44
19	ACTION FORM	+ .15
20	MESSAGE FORM	+ .35
21	THOUGH FORM	+ .30
22	AROUSAL	+ .28
23	URGE	+ .31
24	AFFECTION	+ .25

25	PLEASURE	+ .75
26	DISTRESS	+ .33
27	ANGER	+ .66
28	SENSE	+ .12
29	THINK	+ .24
35	COMMUNICATE	+ .25
36	APPROACH	+ .42
37	GUIDE	+ .26
38	FOLLOW	+ .28
40	AVOID	+ .35
41	ATTEMPT	+ .40

p = 0.01 if r = 0.22; p = 0.001 if r = 0.28 for N = 130

As this kind of data of this kind is quite rare in process research literature, a few comments my be helpful:

For the majority of categories - not only those listed here as examples - we note a statistically significant correlation over the course of treatment. Though statistical significance with an N of 130 is easy to achieve, at least our hypothesis had been that only a few of the categories would reach that level. The results point out that the dyadic relationship organizes the shared "speech world" in such a way that communality is forced onto both speakers (Laffal 1967).

A very high positive correlation occurs with a few categories like PLEASURE and ANGER. From the word lists that make up these categories we can infer that the dialogue around angry and pleasurable topics was very thight and mutually stipulating.

The category DISTRESS referring to all the topics of complaints the level of synchronization did not surpass the average level of all categories. Clinically this could mean that the analyst did not invariably tune in when the patient would complain about nausea, pain and other physical distress symptoms; he would only pick it up as much as necessary to keep in touch.

Up to now there are still few studies with regard to vocabulary or category synchronization; we have followed up this issue in a study of vocabulary similarity with successful and unsuccessful patients (Hölzer et al. 1996).

From this set of data - 61 categories from the adopted Harvard III dictionary for patient and analyst one each - we selected those 55 sessions that were judged by the clinical raters as described in study 1. The category frequencies were dichotomized at the mean of each category individually and entered as independent variables X1.....Xn in a stepwise proceeding multiple regression analysis (BMD O2R, Dixon, 1973). The

dependent variable Y1, Y2 and Y3 were the factor-scores of the factorised clinical judgments as described above.

The model of the multiple linear regression states that the dependent variable Y results from the following linear combination of the independent variables X1 in the following way:

$Y = a1\ x1 + a2\ x2 + ...anXn + C$

The values a1 represent weight factors with which the X1 values are multiplied in order to achieve the maximum of prediction. C represent an additive constant. If Y and X1 are given, then the coefficient a1 have to be selected in such a way that the Non-explanability from Y out of X1 becomes a minimum. This is a problem of differential equations (Gaensslen and Schubö, 1973).

The results of such a stepwise procedure can be described as a pure statistics for the relationship of the content categories and the clinical concepts. If one claims that the sample of 55 sessions is a representative sample for all sessions then the achieved multiple correlation coefficient has to be tested for chance effects.

By performing various pre-tests we singled out of the 2 x 61 variables those variables that presumably would have a high predictive relevance and entered the final analysis with 24 patient-categories and 23 analyst-categories.

Table 5 List of content categories for the final multiple regression analysis

patient variables	mean	analyst variables	mean
FEMALE ROLE	.36	MALE ROLE	.27
NATURAL WORLD	.33	JOBE ROLE	.15
DEVIATION	.38	BODY PART	.38
MESSAGE FORM	.35	TOOL	.33
THOUGHT FORM	.38	TIME	.38
AROUSAL	.44	SOCIAL PLACE	.31
URGE	.31	NATURAL WORLD	.33
PLEASURE	.33	DEVIATION	.11
DISTRESS	.51	THOUGHT FORM	.45
ANGER	.31	AROUSAL	.42
SENSEK	.35	AFFECTION	.23
NOT	.33	URGE	.38
GOOD	.11	DISTRESS	.43
COMMUNICATE	.31	ANGER	.27
GUIDE	.25	THINK	.42
ATTACK	.38	IF	.13
EXPELL	.13	NOT	.13
HIGH STATUS	.33	DEFENSE	.13

PEER STATUS	.33	APPROACH	.25
LOW STATUS	.31	ATTACK	.31
OVERSTATE	.22	MOVE	.29
SIGN STRONG	.29	PEER STATUS	.20
SIGN AFFECT	.31	UNDERSTATE	.24
SEX SENSE	.31		

For each category we also have put down the mean value. As these are dichotomied frequencies, the mean value shows percentage of session, in which the category occurs: 1 = categoy occurs in all session; 0 = category occurs in zero sessions.

Results

1. Predicting of the clinical concept "castration-anxiety"

The stepwise regression resulted in a combination of 15 categories 8 patient and 7 analyst categories that showed a multiple r of + 0.92 with the concept "castration anxiety". The weight of these 15 categories is variable and their direction too (Table 6.).

Table 6: Predicting categories for castration anxiety and their coefficient of regression

patient variables	coefficient	analyst variables	coefficient
FEMALE ROLE	+ 1.44	DISTRESS	+ 0.92
SIGNT AFFECT	+ 1.35	UNDERSTATE	+ 0.80
DEVIATION	+ 0.80	MOVE	- 0.43
COMMUNICATE	+ 0.75	THOUGHT FORM	- 0.55
SEX THEME	+ 0.66	THINK	- 0.78
ANGER	- 0.16	DEFENSE	- 0.96
MESSAGE FORM	- 0.94	AFFECTION	- 1.70
GUIDE	- 1.12		

The content of the predictive categories of the patient refers especially to the female roles and hints at affective states, as they are coded by words as "desire", "mistrust" and "praise". The positive weight of the category DEVIATION, coding the verbal references to his cardiac symptomatology, fits well in the clinical understanding of the patient. The category SEX SENSE fits well in the picture as it refers to sexual topics especially to masturbation and the anxieties of the patient to have damaged himself by this. In the analyst's list of predictive categories we find DISTRESS, which contains the most frequent noun of this analysis "anxiety". We should note it is not the patient's

verbalized anxiety, but the analyst's translation of bodily distress symptoms into anxiety which is predictive in a positive direction.

2. Prediction of the clinical concept "separation-anxiety" with positive transference

The stepwise multiple regression led to a combination of 16 categories (9 patient and 7 analyst) that predicted the clinical concept with a multiple r = + .94.

(a similar discussion follows as under 1.)

3. Prediction of the clinical concept "diffuse anxiety with negative transference"

The stepwise regression led to a combination of 7 categories (3 patient and 4 analyst) that predicted the clinical concept with a r = + 0.77.

(a similar discussion follows)

Clinical discussion of the results

The validity of the impressive predictive power of those combination of content categories is dependent on the validity of the clinical judgments. The basis assumption of the study is that clinicians when making clinical judgments on the basis of reading verbatim protocols base their judgments mainly on verbal elements of the text. In another, more experimental study we could show that short samples of text material that were coded in terms of four anxiety themes by many clinicians could be differentiated reliably by discriminant analysis of computer content categories (Grünzig and Kächele, 1978). If those clinicians change the verbal referents in the process of their judgment procedure unwittingly then the method is bound to procedure artifacts that will not stand any replication. Another problem of the method so far is that the combination of categories might be idiosyncratic for this patient only and clinicians intuitively find their specifics in every case history.

Study 3

With this impressive correlations at hand we set out to try a description of the whole treatment process by using the larger, in fact more than double size sample that we have available.
The empirically found combination of computer scored categories now is used for the production of scores for the clinical concept 1 - 3 over 130 hours, evenly dispersed over the course of treatment. This has been done and the following graphs show the result of this description.
(graphs are available)

The next step in validating these computer based descriptions is to ask a new set of clinicians to judge high and low scoring sessions for the clinical concept.

Bellak L & Smith M B (1956): An experimental exploration of the psychoanalytic process. Psychoanalytic Quarterly, 25, 385-414.
Gottschalk LA, Gleser G C (1969a): The measurement of psychological states through the content analysis of verbal behaviour. Berkeley, University of California Press
Gottschalk LA, Winget C N & Gleser G C (1969b): Manual of instructions for using the Gottschalk-Gleser-Content Analysis Scales. Berkeley, Los Angeles, University of California Press
Grünzig, Hans-Joachim, Horst Kächele und Helmut Thomä (1978): Zur klinisch formalisierten Beurteilung von Angst, Übertragung und Arbeitsbeziehung. Medizinische Psychologie, 4, 138-152.
Hölzer, M, Mergenthaler E, Pokorny D, Kächele H & Luborsky L (1996): Vocabulary measures for the evaluation of therapy outcome: Re-studying transcripts from the Penn Psychotherapy Project (PPP). Psychother Res, 6 (2), 95-108.
Laffal, J (1967): Characteristics of the three-person conversation. *J Verb Learn Verb Behav*, 6, 555-559.
Stone, P. J., D. C. Dunphy, M. S. Smith und D. M. Ogilvie (1966): The General Inquirer: a computer approach to content analysis. Cambridge, Mass., M.I.T Press
Strupp, Hans H., Jacob B. Chassan und J A Ewing (1966): Toward the longitudinal study of the psychotherapeutic process. In: Gottschalk, L A Auerbach A H (Eds) Methods of research in psychotherapy. New York Appleton-Century-Crofts, S. 361-400