Vocabulary measures for the evaluation of therapy outcome: studying the transcripts from the Penn Psychotherapy Project (PPP)

M. Hölzer, H. Kächele, E. Mergenthaler University of Ulm &
L. Luborsky (Penn Medical School)

Introduction

In the Penn Psychotherapy Project (PPP) Luborsky et al. (1980) have identified two types of patients' experiences of being in a helpful relationship with a therapist: The Helping Alliance type I and type II, which were regulary observed in therapies with favorable outcomes. Probably the most important finding of the Penn Project was that such an experience accounts for more outcome variance than most of the pretreatment variables examined. Thus it could be concluded that predicting outcome is more successful with in-treatment information about the communication between therapist and patient rather than patient and therapist variables that leave out the actual therapeutic interaction (see also Luborsky et al. 1988).

The guiding rationale for the work we report here, is that successful psychotherapy and the concomittant experience of a "helping alliance" should in some way be reflected in characteristics of both patient and therapist vocabulary.

Since computer-aided text analysis has been the central interest of the Ulm Psychotherapy Department (Kächele & Mergenthaler 1983; Mergenthaler & Kächele 1990), collaboration with Luborsky and working with material

1

from the PPP provided a further incentive, i.e., the incorporation of an English language section into the Ulm Textbank System (Mergenthaler & Kächele 1988).

Before we present the methods and preliminary findings, we offer brief description of the PPP sample examined.

- Here insert table 1 -

The Penn Project has led to the identification of the ten most and ten least improved patients who were treated for at least 25 sessions and selected on the basis of two outcome measures from a total sample of 73 cases (one outcome measure on the table). The median length of treatment was 61 weeks for improvers, and 43 weeks for non-improvers. The 20 patients were treated by 18 psychodynamically oriented, male therapists, ten of them second or third-year residents under supervision and eight of them more experienced. There were no significant differences between the most versus the least improved cases with respect to their demographic characteristics. According to the DSM II all 20 were nonpsychotic outpatients, although they differed considerably in their diagnoses (e.g. adjustment reaction of adolescence, obsessive-compulsive personality disorder, depressive reaction and schizoid personality).

The therapy sessions were audiotaped and transcribed. For each patient four segments were drawn, two from the initial as well as two from the final stage of treatment, resulting in a total of 80 segments. Each segment consisted of the first 30 minutes of a session. In order to facilitate the evaluation, the early-in treatment segments were combined, as were the late-in treatment segments, yielding two "texts" for each patient, one for each stage of therapy.

From the various computer measures studied, three of increasing complexity will be presented here.

1. verbal activity
2. vocabulary measures
3. content-analytic measure

1. Verbal Activity

In the third edition of the Handbook of Psychotherapy and Behavior Change (Garfield & Bergin 1986), Orlinsky and Howard state that "if the patient doesn't talk much of the time and the therapist doesn't talk some of the time, then it would be hard to claim that psychotherapy is actually taking place." In spite of its objectivity and the ease, with which it is obtained, verbal activity is often seen as too simple a measure to reflect the complexity of what is going on in psychotherapy and thus viewed as theoretically unimportant. But this attitude among psychotherapy researchers might reflect their prejudice against simplicity, rather than empirically supported facts. There is still not enough well established knowledge about how the amount of patient's and therapist's speech relate to each other and how that influences psychotherapeutic processes. A considerable number of studies have shown that verbal activity is significantly linked to outcome. Particulary, "a majority of findings (7 of 11) relating the amount of patient speech to outcome were significantly positive." In studies of therapist speech only 5 of 14 showed significantly positive findings "while 8 showed no relationship of therapist talkativeness to patient benefit." However, the "balancing" of patient and therapist conversational activity, measured by the ratio of therapist to patient verbal activity, might be a more potent predictor of outcome than simple activity measures. Furthermore when studying the phenomena endeatil, wiede variations of verbal activity may be observed especialy in longer treatments (Kächele 1983).

2. Vocabulary measures

As opposed to verbal activity, formal vocabulary measures like TTR have been tried only in the early phases of psychotherapy research and don't belong to the present battery of psychotherapeutic research tools, although they might help fill the gap between formal and content related approaches. As opposed to Verbal Activity which is defined as the total number of words (Tokens) occurring in a given text, the term "vocabulary" always refers to the number of different words (Types) that are used by a speaker. The ratio between Types and Tokens, the Type-Token-Ratio, has been usually looked at as an indicator of the diversity of a text. From a research perspective, vocabulary measures defined in terms of Types are interesting, not only because they are easily and objectively

obtained. Since words stand for concepts (and therapy has essentially to do with an exchange of concepts and beliefs, with assimilation of new material and accomodation of previous schemata), changes in vocabularies during treatment might parallel or at least partly reflect such learning processes.

In a therapeutic dialogue two kinds of vocabularies can be distinguished:

1.The Private Vocabulary (PV), i.e., the set of Types that are used by only one of the speakers, here denoted as PPV and PTV.

2.The Intersectional Vocabulary (IV), the set of Types that are used by both patient and therapist.

A slightly more sophisticated way to compute the Private Vocabulary results in what we call the "Characteristic Vocabulary". Here the decision as to whether a certain Type belongs to the vocabulary is based on probabilities and is thus a statistical one. A word has to occur in the text of one speaker a number of times as often as in the text of another to be incorporated in his characteristic vocabulary. Depending on the chosen probability, the magnitude of the Characteristic Vocabulary may differ considerably.

Since vocabulary measures have not yet been sufficently validated, up to now their interpretation has been guided only by clinical experience.

Our hypotheses were:

1. The ability of a therapist to accomodate to the language of his or her patient, to bridge social differences and to empathize with the patient should result in a low Private Vocabulary on his part. Therapists of improvers are likely to show more of these capacities than therapists of non-improving patients.

2. The ability of patients to process or integrate the concepts and beliefs of the therapist might be reflected in a higher joint vocabulary at the end of therapy than in early stages. Again, this effect should be stronger with improvers than with non-improvers.

3. The RID

As a third evaluation procedure we applied the RID (the Regressive Imagery Dictionary as developed by C. Martindale (1975) for use in computerized content analysis) to the transcripts. Psychoanalytic theory holds, that in order for change to occur in the patient, regression towards primary process thinking must be achieved. Free association, the "basic rule" of psychoanalysis is generally viewed as inducing such a movement from secondary process cognition to the more primitive and less mature primary process. Although in the light of recent findings of cognitive psychology the theory of the "regression in the service of the ego" (Kris, 1952) needs some revision, the categories with which Martindale's Regressive Imagery Dictionary measures primary and secondary process phenomena are nevertheless interesting: "Primary process is based upon the occurrence of words with connotations of Drives, Sensations, Defensive Symbolization (i.e., references to disorder in the external world), Regressive Cognition (direct references to alteration in states of consciousness), and Icarian Imagery (i.e., references to fire, water, rising and falling). Secondary process tabulates references to abstraction, instrumental behavior, social behavior, time, moral imperatives, order and restraint" (Reynes, et al., 1984). In its original version the RID contains a total of about 2900 words, which are assigned to 43 exclusive categories. The Ulm version of the RID works with a total number of more than 5000 words, reduced their basic forms.

- Here insert table 2 -

Categories 1 to 29 make up the primary process, while categories 30 to 36, yield the secondary process. A third subset of categories (37 to 43), labeled as "emotions", were also assessed in the transcripts. They contain words refering to positive and negative affects such as anxiety, sadness, anger, etc.

The Regressive Imagery Dictionary has shown good construct validity in studies of a wide variety of texts. As Reynes reports: "More primary process and less secondary process imagery have been found in folk tales of primitive as opposed to complex preliterate societies, in poetry of writers who exhibit signs of psychopathology as opposed to writers who do not show such signs and in stories told by younger as opposed to older

5

children."

Reynes, Martindale and Dahl (1984), evaluating working, neutral and resistance sessions of a psychoanalytic therapy by means of the RID, found that primary process increases and secondary process decreases significantly as one moves from resistance through neutral to working sessions. Because patient and therapist "did not exhibit radically different amounts of primary process in their speech," Reynes et al. concluded that "psychoanalytic work involves a movement toward primary process cognition on the part of both patient and analyst." Although the PPP patients in the present study were treated by psychoanalytically informed psychotherapy, we hypothesized that success in supportive-expressive therapy as outlined by Luborsky (1984) should also positively correlate with the primary process and the emotion subset of categories as well.

Results

- Here insert table 3 -

1. Verbal Activity
Although both groups show considerable within-groups variability, patients and therapists of the improving and the non-improving group, don't exhibit any differences in their overall Verbal Activity, i.e., in their average scores over all 4 sessions. In this respect their conversational behavior seems to be almost identical. Assuming that a therapeutic session lasts approximately 50 minutes, patients utter about 5700 and therapists 1600 words per session. For both groups the mean ratio of therapist talkativeness to patient talkativeness was 1/3.5, meaning that patients in general talk three times as much as their therapists. Thus, we could not confirm Scobel's (1979) results who had found a 1 to 2 ratio for successful and a 1 to 3 ratio for non-successful Rogerian psychotherapy. Still, an interesting and significant difference between improvers and non-improvers was found when we looked at this ratio at the different points of time in treatment.

- Here insert table 4 -

Moving from the initial to the final stage in therapy the ratio drops with

improvers and rises with non-improvers, meaning that therapists of improvers talk relatively less and therapists of non-improvers talk relatively more in relationship to their patient at the end of the treatment compared with the beginning. Clearly, this hints at an interaction effect of time and outcome with respect to relative Verbal Activity and has should be further examined.

In addition to this, we found two significant negative correlations between Verbal Activity and Residual Gain, one of the outcome measures applied.

- Here insert table 5 -

A verbally active therapist seems to have a negative impact on improving patients, whereas more talkative patients among the non-improvers tend to show the least improvement of this group.

2. Vocabulary measures

Although not as impressive as in Verbal Activity, a considerable variation in their Private and Intersectional Vocabularies is seen in both therapy groups.

Comparing the vocabulary measures between improvers and non-improvers does not lead to significant differences, with one exception.

- Here insert table 6 -

The correlation between Residual Gain as outcome measure and the size of Private Patient (PPV) Vocabulary distinguishes improvers (they show a mildly positive correlation) and non-improvers, with whom we found a significantly negative correlation. Again, this indicates an interaction effect. The Private Patient Vocabulary of improvers might well reflect personal attitudes and qualities of the patient, such as relative autonomy or certain cognitive capaciies, whereas large Private Patient Vocabulary with non-improvers may hint to a rather monologue style of conversation, the content of which was not fully processed and integrated by the therapist. We could not support our hypothesis of a lower Private Therapist Vocabulary (PTV) among improvers. Furthermore, in both groups the Private Vocabularies

stay essentially the same throughout the treatment; there are no differences discernable between the initial and the end phase of therapy.

Another interesting finding is the significantly negative correlation between the therapist's experience and his Private Vocabulary, which was found in both groups. Here perhaps it is fair to say that experienced therapists tend to not distance themselves from the lexical capacities of their patients.

The Intersectional Vocabulary between patient and therapist exhibits a significant decrease with improvers: From a mean of 241 for the early-in-treatment text it drops to 207 words for the late-in-treatment segments (t = 2.32, p<.05) Obviously, this finding runs counter to our second hypothesis, in which we postulated the reverse relationship, viewing an increase in the Intersectional Vocabulary as an indicator for a more successful therapeutic exchange. Here, our knowledge about the Intersectional Vocabulary is not yet sufficient to formulate valid hypotheses. This vocabulary seems to be more complicated and composed of a variety of sub-vocabularies, i.e., on the one hand words that are necessary to construct meaningful language, and on the other hand, words that refer to specific and maybe therapeutically important content categories. Our next step will be to determine these sub-vocabularies and to formulate differential hypotheses.

3. Regressive Imagery Dictionary (RID)

Evaluation of the transcripts by the RID yielded the following ratio of distribution.

- Here insert table 7 -

While more than 80% of the text was not coded by this dictionary, both patients and therapists do not exhibit significantly different amounts of primary and secondary process in their speech. This is true for both groups, improvers and non-improvers, who also don't show significant differences in the emotion words. An average of 2.1% primary process of the total text is a rather low percentage compared with findings of other studies.

Here the percentage varies from 2.9 to 6.9%. The 1/5 ratio of primary to

secondary process found in our study compares best with the resistance hours in Reynes et al.'s investigation.

The concept "sensation", covering single categories like touch, taste, odor, sound or vision contributes significantly more to the primary process score than the other concepts, both in patient and in therapist speech.

- Here insert table 8 -

Furthermore, in both groups, therapists utter relatively more words listed in the sensation categories than their patients. This finding was highly significant. Because the sensation concept does not discriminate poor from successful outcome, this result may hint at a general therapeutic attitude or stance. It can be interpreted in the light of the Dual Code Theory and Bucci's (1985) findings about Referential Activity, which refer to two different representational systems, a verbal and a nonverbal one, operating in the mind. In this theory, Referential Activities are seen as linking or bridging functions between the two systems and manifested in concrete and specific language and by the use of words which refer to perceptual or sensory input. Maybe therapists, by prefering a rather concrete way of talking and thus showing high Referential Activity themselves, are trying to induce Referential Activity on part of their patients as well. We are planning to test this hypothesis by correlating the RID measures with ratings of Referential Activity.

In conclusion, undoubtedly the linguistic analysis of the Penn-transcripts is made more difficult by the wide variety of diagnoses found in the 20 cases examined. Clearly, improvement and the way in which it is reflected lingustically in a case of 'latent schizophrenia' is different from improvment in a case of adjustment reaction. Therefore, the measures mentioned above -were not expected to fully explain the 'linguistic variance' inherent in our sample.

Refinement of the methods applied seems to be necessary on two levels: .
1. The vocabulary methods clearly need revision with respect to the existence of differently relevant sub-vocabularies and related differential hypotheses.

2. Content analytic methods as the RID have to be further validated.

Although our evaluation of the Penn-transcripts by means of computer aided methods is still in an early stage, we are optimistic that these methods will contribute to the nomological net with which psychotherapeutic process and outcome might be explainable on the level of microprocesses.

Bibliography

Bucci, W (1985). Dual Coding. A Cognitive Model for Psychoanalytic Research. Journal of the American Psychoanalytic Ass., 33:, 571-607.

Kächele H (1983) Verbal activity level of therapists in initial interviews and long-term psychoanalysis. In: Minsel WH W (Hrsg) Methodology in psychotherapy research. Lang, Frankfurt, S 125-129

Kächele H, Mergenthaler E (1983) Computer-aided analysis of psycho-therapeutic discourse. In: Minsel R, Herff W (Eds) Methodology in Psychotherapy Research. Proceedings of the 1st European Conference on Psychotherapy Research, Vol. 1. Peter Lang, Frankfurt, S 116-161

Kris, E. 1952. Psychoanalytic Explorations in Art. Int. University Press, New York.

Luborsky L (1984) Principles of psychoanalytic psychotherapy. A manual for supportive-expressive treatment. Basic Books, New York.

Luborky, L., Mintz, J., Auerbach, A. et al. 1980. Predicting the outcome of psychotherapy. findings of the Penn Psychotherapy Project. Arch. Gen. Psychiatry, 37:471-481.

Luborsky L et al (1988) Who benefits from Psychotherapy? Basic Books, New York

Martindale, C. (1975). Romantic Progression. The Psychology of Literary History. Hemisphere, Washington.

Mergenthaler E, Kächele H (1988) The Ulm Textbank management system: A tool for psychotherapy research. In: Dahl H, Kächele H, Thomä H (Hrsg) Psychoanalytic process research strategies. Springer, Berlin Heidelberg New York London Paris Tokyo, S 195-212

Mergenthaler E, Kächele H (1990) The Ulm Textbank Research Program. In: Beutler (Hrsg) International Psychotherapy Research Programs. Pergamon, New York, London, S

Mowrer OH (1953) (Ed) Psychotherapy, Theory and Research. Ronald Press, New York

Orlinsky, D.E., Howard, K.I. 1986. The Relation of Process to Outcome. In

Garfield, S.L., Bergin, A.E. (Eds).: Handbook of Psychotherapy and Behavior Change (Third Edition). Wiley, New York.

Reynes, R., Martindale, C., Dahl, H. 1984. Lexical Differences between Working and Resistance Sessions in Psychoanalysis. Journal of Clinical Psychology, 40:733-737.

Scobel, W. A. 1979. Sprache und Gesprächspsychotherapie. Eine inhaltsanalytische Untersuchung. Psychothera. med. Psychol., 29:66-72.