An automated method of content analysis for psychotherapy research: A further validation

Sergio Salvatore, Omar Carlo Gioacchino Gelo, Alessandro Gennaro, Roberto Metrangolo, Grazia Terrone, Valeria Pace, Claudia Venuleo, Annalisa Venezia & Enrico Ciavolino

To cite this article: Sergio Salvatore, Omar Carlo Gioacchino Gelo, Alessandro Gennaro, Roberto Metrangolo, Grazia Terrone, Valeria Pace, Claudia Venuleo, Annalisa Venezia & Enrico Ciavolino (2017) An automated method of content analysis for psychotherapy research: A further validation, Psychotherapy Research, 27:1, 38-50, DOI: 10.1080/10503307.2015.1072282

To link to this article: http://dx.doi.org/10.1080/10503307.2015.1072282

Published online: 04 Sep 2015.

Article views: 115

View related articles

View Crossmark data

Citing articles: 1 View citing articles
An automated method of content analysis for psychotherapy research: A further validation

SERGIO SALVATORE1,2, OMAR CARLO GIOACCHINO GELO1,3, ALESSANDRO GENNARO1,2, ROBERTO METRANGOLO2, GRAZIA TERRONE4, VALERIA PACE2, CLAUDIA VENULEO1,2, ANNALISA VENEZIA1, & ENRICO CIAVOLINO1

1Department of History, Society and Human Studies, University of Salento, Lecce, Italy; 2Centro Studi di Psicoterapia e Psicologia della Salute (CESP)—Istituto Scientifico Biomedico Euro Mediterraneo (ISBEM), Lecce, Italy; 3Department of Psychotherapy Science, Sigmund Freud University, Vienna, Austria & 4Department of Human Studies, University of Foggia, Foggia, Italy

(Received 14 February 2014; revised 22 June 2015; accepted 1 July 2015)

Abstract

Objective: The aim of the study is to validate the ability of ACASM (Automated Co-occurrence Analysis for Semantic Mapping) to provide a representation of the content of the therapeutic exchange that is useful for clinical analysis.

Method: We compared the clinical case analyses of a good outcome psychodynamic therapy performed by a group of clinicians (n = 5) based on the verbatim transcripts (transcript-based analysis) with the clinical case analyses performed by another group of clinicians (n = 5) based on the ACASM representation of the same sessions (ACASM-based analysis). Comparison concerned two levels: the descriptive level and the interpretative level of the clinical case analysis. Results: Findings showed that, inconsistently with our hypothesis, ACASM-based descriptions of the case obtained worse evaluations than transcript-based descriptions of the case (on all 3 criteria adopted). On the contrary, consistently with our hypothesis, ACASM is undistinguishable from the verbatim transcripts as regards the case interpretation (on 2 out of 3 criteria adopted).

Conclusions: ACASM provides a description of the case that, though different from the one provided by the transcripts, enables clinicians to elaborate clinical interpretations of the case which approximate those produced by clinicians working directly on verbatim transcripts.

Keywords: ACASM; automated text analysis; thematic analysis; psychotherapy process

Psychotherapy may be regarded as a communicational exchange between clients and therapists that unfolds over sessions. Within this exchange, the content of what is said (i.e., the semantic level of communication) plays a primary—although not exclusive—role. It follows that text analysis may represent a very useful tool for psychotherapy process researchers (Salvatore, Gennaro, Auletta, Tonti, & Nitti, 2012).

Text analysis is generally defined here as any procedure of inquiry of a text used to draw meaningful inferences on its content (Gelo, Salcuni, & Colli, 2012). Most of the methodologies and related procedures (henceforth, for the sake of brevity: methods) of text analysis currently used in psychotherapy process research are usually performed by human raters and/or judges (Gelo & Manzo, 2015). Examples of these are transcript-based rating scales and/or category systems, as for example the core conflictual relationship themes (CCRTs) (Luborsky, Popp, Luborsky, & Mark, 1994), the structural analysis of social behavior (Benjamin, Rothweiler, & Critchfield, 2006), the comprehensive psychotherapeutic interventions rating scale (Trijsburg, 2002), the collaborative interactions scale (Colli & Lingiardi, 2009), the metacognition assessment scale (Semerari et al., 2003), and the grid of the models of interpretation...
procedures of qualitative data analysis (Glaser & Strauss, 1967; Strauss & Corbin, 1998), phenomenological analysis (Giorgi, 2009), narrative analysis (Avdi & Georgaca, 2007), conversation analysis (Madill, Widdicombe, & Barkham, 2001), consensus qualitative research (Hill et al., 2005), thematic and content analysis (Braun & Clarke, 2006); or procedures of data analysis combining qualitative and quantitative approaches, such as for example task analysis (Greenberg, 2007), comprehensive process analysis (Elliott, Fischer, & Rennie, 1999; Elliott, 1989), and assimilation analysis (Stiles et al., 1990) (see Gelo et al., 2012 for discussion). These different methods have been successfully applied in psychotherapy process research; nonetheless, one of their major shortcomings is the fact that they are extremely time- and energy-consuming, especially because of the efforts which must be made to ensure adequate levels of validity, reliability, and/or credibility (see Elliott, Fischer, & Rennie, 1999; Hill & Lambert, 2004). For these reasons, some automated methods of text analysis have been developed and applied in recent years in the field of psychotherapy process research, such as the therapeutic cycle model (Mergenthaler, 1996, 2008) and referential activity (Bucci, W., Kabasakalian, R., & The RA Research Group, 1992). These methods follow a top-down approach to textual analysis, in which predefined categories are used to categorize units of text. Each of these categories is usually called the dictionary, which contains all the words considered to be indicative of the content represented by that category. Although these methods of top-down textual analysis have been shown to allow a reliable and valid investigation of the therapeutic process, they present a major limitation: They disregard the contextual nature of linguistic meaning (Salvatore et al., 2012). To say that meaning is contextual means acknowledging that the meaning of any word is not fixed but depends on the way it combines with other words in the contingent dynamics of talk. In other words, meaning depends on the contextual connections it establishes with other words, and cannot thus be adequately depicted by the application of context-blind rules of coding. Based on these considerations, there has recently been a call for the development of automated procedures of content analysis which, thanks to their bottom-up logic, allow context-sensitive text analysis (Salvatore et al., 2012).

**Automated Co-occurrence Analysis for Semantic Mapping**

In order to overcome the limitations of the traditional automated methods of text analysis employed in psychotherapy research described above, the Automated Co-occurrence Analysis for Semantic Mapping (ACASM) has been recently developed (Salvatore et al., 2012). ACASM is a bottom-up procedure of text analysis “based on explicit, invariant rules of coding and yet able to take the contextuality of meaning into account” (Salvatore et al., 2012, p. 258; italics added). ACASM is aimed at extrapolating thematic contents (i.e., semantic meanings) active in the text. Each thematic content is characterized by a cluster of words which tend to co-occur (i.e., to be associated with each other) throughout the text. Such sets of co-occurring words are identified by means of invariant but context-sensitive computational rules automatically implemented by an ad-hoc software. ACASM is a member of the broader family of methods of semantic analysis focusing on the co-occurrence of lexical units (e.g., ALCESTE, cf. Reinert, 1993; latent semantic analysis, cf. Landauer & Dumais, 1997; meaning extraction method (MEM), cf. Chung & Pennebaker, 2008; see Wolf, Chung, & Kordy, 2010 for an application of MEM within the clinical context). Compared to most such methods, the main specificity of ACASM is that it adopts a single sentence or group of a few sentences as its unit of context (the unit of context is the segment of text within which co-occurrences are detected). This unit of context is narrower than the one adopted by most other methods. ACASM chooses this unit of context in order to make the semantic analysis sensitive to the contingencies of the communication—namely, how words tend to be combined with each other in a given circumscribed moment of time. The contingency of communication is an important component of the contextuality of meaning (Salvatore, 2015), playing a central role in particular in the case of spoken communication, as the clinical exchange is.

A first study of ACASM validity (Salvatore et al., 2012) was conducted in order to verify the assumption that ACASM is functionally equivalent to a model of human bottom-up semantic analysis based on commonsense (Garfinkel, 1967; Valsiner, 2007). To this end, a *Turing-like* criterion of validity was adopted, namely, it was expected that “ACASM could be considered a valid method of semantic analysis if and only if the analysis it produces cannot be distinguished from those produced by expert human coders” (Salvatore et al., 2012, p. 258). Such a criterion was adopted because “in the case of bottom-up semantic analysis it is not possible to refer to an external, objective normative criterion in accordance with which to evaluate the validity of the analysis in absolute terms” (Salvatore et al., 2012, p. 258). Thus, the performances, respectively, of ACASM and of a group of human
coders (blind to the aims of the study) were compared with regard to two specific tasks of semantic analysis: (i) the evaluation of semantic similarity among different text units and (ii) the bottom-up classification of a certain amount of these text units into clusters (i.e., thematic contents) based on their semantic meaning. Results showed that the performance of ACASM is indistinguishable from that of human coders with regard to these two tasks, thus providing initial empirical evidence to the validity of semantic analysis performed by ACASM.

The study mentioned above showed that ACASM is able to perform two operations of semantic analysis (namely, evaluation of semantic similarity of text-units and their bottom-up classification according to their thematic content) in a way that is indistinguishable from human raters, with the advantage of being more reliable and time-saving. However, it does not tell us anything regarding the ability of ACASM to provide a representation of the textual content useful for clinical analysis. In this regard, in fact, Salvatore et al. (2012) concluded their paper by stressing the need to verify whether the semantic analysis provided by ACASM may be usable for clinical purposes.

**Aims of the study**

The present study aims at further validating ACASM in this direction, namely, with regard to its ability in supporting human judges in making meaningful inferences on a clinical case (i.e., clinical case analysis). Within the context of this study, a clinical case analysis is defined as the ability of a subject to: (1) Identify recurrent themes within a clinical case and depict their temporal evolution throughout it (case description); (2) Provide a clinical interpretation of the case based on 1 (case interpretation).

Our general hypothesis is that the clinical case analysis (in terms of points 1 and 2 described above) based on the verbatim transcripts of a case (transcript-based analysis) is indistinguishable from the clinical case analysis based on ACASM content analysis of the same transcripts (ACASM-based analysis). In order to verify this, we test the null hypothesis that there are no differences in the clinical cases analyses—at the level of both (1) case description and (2) case interpretation—performed by expert clinicians by means of either a transcript-based or an ACASM-based analysis.

**Method**

**Sample and Material**

The sample in this study were N = 10 experienced psychotherapists and clinical psychologists with a psychodynamic orientation (2 males, 8 females; mean age = 35.7, SD = 5.3). They were recruited among a network of private practitioners established by the University of Salento (Italy) and by the Sigmund Freud University (Austria); they were asked to participate in a study involving the analysis of a case of psychotherapy.

Participants were assigned randomly to experimental (EXP) and control (CONTR) groups. The EXP group was composed of all female psychotherapists (mean age = 36.4, SD = 6.2). The CONTR group was made up of 2 males and 3 females (mean age = 35, SD = 4.9). No difference was found as to age. No statistical difference was found as to the sex (chi-square; exact Fischer test) either, but in this case this should have reflected the low power of the comparison.

The psychodynamic psychotherapy of Max was used as the data source. Max is an Italian man, at the beginning of the treatment, being a 35-year-old manager diagnosed with a Schizoid Personality Disorder (DSM-IV-TR; American Psychiatric Association [APA], 2000) who requested therapy because of relational problems, particularly with his wife and his family of origin. The therapy was conducted by an experienced clinician with a psychodynamic background on a weekly basis and resulted in 74 sessions over 2 years. The treatment had a good outcome according to: (a) the clinical judgment of the therapist, which was confirmed in two follow-up sessions, respectively, at 4 and 8 months after the termination of the treatment; (b) the pre-post application of the Shedler-Westen Assessment Procedure-200 (SWAP-200; Westen & Shedler, 1999a, 1999b) and of the CCRT (Luborsky et al., 1994). The SWAP-200 showed a decrease below the clinical relevance threshold of the personality disorder factor measuring schizoid, obsessive, and avoidant traits. The CCRT highlighted a major change in “Response from other” and “Response from the Self” components. Such a difference can be interpreted as indicative of the fact that while at the beginning of the clinical process the patient perceived other people as opposed to him, and felt depressed and ashamed; at the end of the clinical process he felt he was appreciated and respected by others as well as able to accept himself (see Rocco, De Bei, & Mariani, 2013 for more details).

All the sessions were recorded after informed consent was obtained from the client and were transcribed according to international standards used in psychotherapy (Mergenthaler & Stinson, 1992), adapted for the Italian language (Mergenthaler, Freni, Giampieri, & Ferrari, 1998).

**Design and Procedure**

The study was organized in two stages: The first aimed at obtaining a clinical case analysis from each
participant; the second aimed at comparing the clinical case analyses obtained.

Stage 1. Clinical case analysis. First, we asked each psychotherapist of both groups (i.e., EXP and CONTR groups) to provide a clinical case analysis of Max’s psychotherapy. One group (the EXP group) had to base its clinical analysis on the case representation consisting of the output produced by ACASM’s semantic analysis of transcripts of the sessions comprising the case under investigation (i.e., ACASM-based clinical case analysis); the other group of clinicians (the CONTR group) had to rely on the case representation consisting of the verbatim transcripts of the sessions (i.e., transcript-based clinical case analysis). Thus, ACASM processed the same textual corpus that was read by CONTR group clinicians, consisting of the verbatim transcript of all Max’s psychotherapy. This corpus comprised 422,703 words.

ACASM-based analysis. The verbatim transcribed sessions of the case under investigation were first subjected to ACASM analysis. To this end, the following operations were performed (see Salvatore, Gennaro, Auletta, Tonti, & Nitti, 2010 for more details on the procedure): First, each session was automatically segmented into elementary context units (ECUs) according to the following criteria: (i) Each ECU begins just after the end of the previous ECU; (ii) Each ECU ends with the first punctuation mark (“.”, “;”, “!”, or “?”) occurring after the threshold of 250 characters from the first character; (iii) If an ECU is longer than 500 characters, it ends with the last word found within such a length, even if there is no punctuation mark.

Second, a dictionary for the text-analysis was constructed. In order to do this, each lexical form present in the transcripts of all the sessions comprising the considered case was categorized into the lemma it belongs to, thus producing a list of lemmas present in the textual corpus analyzed. Thereafter, the first 5% of the most frequent lemmas were excluded; this was done since very high-frequency lemmas (such as “to,” “and,” “of,” etc.) tend to co-occur in too many different ECUs, thus reducing their ability to discriminate among different patterns of co-occurrence. Then, 10% of the most frequent lemmas of the remaining list were selected, in order to reduce the variability of the data-set.

Third, the textual corpus was digitally represented in terms of a matrix displaying ECUs in rows and lemmas in columns; the cell $x_{ij}$ received the value “1” if the $j$th lemma was contained in the $i$th ECU, and the value “0” otherwise.

Fourth, a cluster analysis (Aldenderfer & Blashfield, 1984) was applied to the matrix in order to group the ECUs of the textual corpus into clusters, using the co-occurrence of lemmas among the ECUs as a criterion of similarity—in this way, each cluster could be considered a thematic content active in the textual corpus, since it contained groups of utterances (i.e., ECUs) sharing a similar pattern of co-occurring lemmas, therefore semantic content. The number of clusters in which the text is segmented is defined in accordance with an iterative algorithm; the procedure of clustering stops when further partitions produce no further significant improvement of the inter-/intra-cluster ratio, which means that increasing the number of clusters does not produce an appreciable increment of information. (This procedure is analogous to the saturation process in bottom-up qualitative analyses, through which informational redundancy is reached [Morse, 1995; Mörtl & Gelo, 2015].) As a result of the procedure of analysis presented above, 12 clusters were identified.

Finally, each of these clusters was labeled through a process of consensus reaching by the first, second, and third authors of the present study; the latter were blind to the transcript-based analysis performed by the CONTR group (see below). Table I shows the 12 thematic contents identified.

After this procedure was completed, the previously identified themes were displayed on a PowerPoint file together with their temporal evolution over, respectively, 10 and 3 phases.

Both the labeling of themes and the calculation of their distributions over time were part and parcel of the ACASM output (see Figures 1 and 2 for an example).1

Finally, each psychotherapist of the EXP group received the PowerPoint file containing the ACASM output and was asked, after having carefully read the content of the file, to carry out a clinical case analysis consisting of a (1) case description (i.e., identification of the recurrent themes of the therapy and description of their temporal evolution) and a (2) case interpretation (i.e., clinical interpretation of the case based on the case description). The instructions are reported in Table II.

Transcript-based analysis. In the case of the transcript-based analysis, the fully transcribed sessions of the case under investigation were given to the judges of the CONTR group. Their task was to carefully read all the transcripts and then to carry out a clinical case analysis following the same instructions given to the EXP group for the ACASM-based analysis (see Table II).

Stage 2. Comparison of clinical case analyses. In the second step of this study, we asked five more judges (all female; mean age = 28.6, SD = 4.6) to compare the different clinical case analyses (i.e., case description plus case interpretation) previously produced by the members of the EXP group (ACASM-based analysis) and CONTR group (transcript-based
All five judges were psychologists with at least an average level of clinical competence.

In order to compare the ACASM-based and transcription-based analyses, each of these 5 judges received, in random order, all 10 texts comprising the transcription-based and ACASM-based clinical case analyses; judges were blind to the fact that those texts stemmed from two distinct groups working differently. Their task was to evaluate the material with regard to the following set of criteria (see Table III for definitions): (i) case description: descriptive similarity, accuracy, and efficacy of synthesis and (ii) case interpretation: interpretative similarity, appropriateness, and clinical validity.

All the criteria were measured on a 7-point Likert scale (see Table III for details).

As Table III shows, the criteria (b) accuracy, (c) efficacy of synthesis, (e) appropriateness, and (f) clinical validity entail a case-centered evaluation, while (a) descriptive similarity and (d) interpretative similarity...
entail a pair-centered evaluation. The former set of criteria is applied to each single clinical case analysis (e.g., “how accurate is this case?”), while the second set of criteria is applied to pairs of clinical case analyses (e.g., “how coherent are the case descriptions of case x and y?”).

Accordingly, the former type of criteria (case-centered ones) were applied separately to each single clinical case analysis, while the second type of criteria (pair-centered ones) concerned the 45 pairs obtained through the one-to-one combination of the 10 clinical case analyses. Of these, \( n = 10 \) were pairs of exclusively ACASM-based analyses (AA), \( n = 10 \) pairs of exclusively transcript-based analyses (TT), and \( n = 25 \) pairs of a transcript-based and an ACASM-based analysis (AT).

The mean agreement among the five raters was assessed using the intra-class correlation average measure (ICC; Shrout & Fleiss, 1979). This was done separately for each criterion. The level of agreement varied between .79 and .96 (accuracy: ICC = .93; efficacy of synthesis: ICC = .84; appropriateness: ICC = .92; clinical validity: ICC = .96; descriptive similarity: ICC = .79; interpretative similarity: ICC = .84). These figures may be considered as indicative of an “excellent” agreement (Shrout & Fleiss, 1979).
Table III. Criteria and instructions used for the comparison of clinical case analyses.

<table>
<thead>
<tr>
<th>Case description</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>For each case description, please evaluate the extent to which these are similar to each other concerning their content. Evaluation modality: 7-point Likert scale (1 = not at all similar; 7 = completely similar).</td>
</tr>
<tr>
<td><strong>Clinical validity</strong></td>
<td>For each case interpretation, please indicate how much clinical validity it shows. Evaluation modality: 7-point Likert scale (1 = not at all clinically valid; 7 = completely clinically valid).</td>
</tr>
<tr>
<td><strong>Descriptive similarity</strong></td>
<td>For each pair of case descriptions, please evaluate the extent to which these are similar to each other concerning their content. Evaluation modality: 7-point Likert scale (1 = not at all similar; 7 = completely similar).</td>
</tr>
<tr>
<td><strong>Efficacy of synthesis</strong></td>
<td>For each case interpretation, please evaluate the extent to which these are similar to each other concerning their content. Evaluation modality: 7-point Likert scale (1 = not at all similar; 7 = completely similar).</td>
</tr>
<tr>
<td><strong>Interpretative similarity</strong></td>
<td>For each pair of case interpretations, please evaluate the extent to which these are similar to each other concerning their content. Evaluation modality: 7-point Likert scale (1 = not at all similar; 7 = completely similar).</td>
</tr>
</tbody>
</table>

Data Analysis

In order to test our hypothesis, we compared the performances of ACASM-based and transcript-based clinical case analyses. Evaluations concerning the criteria entailing, respectively, case- and pair-centered evaluations were analyzed separately. Concerning the former set of criteria (case-centered criteria: accuracy and clinical validity), we adopted a random intercept model with the 5 raters as random effect and the two groups (i.e., EXP vs. CONT) as fixed effect. Assumption of normality (Shapiro–Wilk test) and homoscedasticity (Levene test) were tested and verified for each case-centered criterion.

We chose to use a random intercept model in order to increase the statistical power of the comparison, thereby making use of the multiple raters as a way for increasing the number of observations. Indeed, due to the fact that we adopted a Turing-like criterion of validity, the desirable output of the study would be the non-disconfirmation of the null hypothesis; consequently, it was necessary to obtain an acceptable level of statistical power in order to make results meaningful. Moreover, consistently with that consideration, we chose to adopt a less conservative level of α (.10). Given such a level of α, and given the number of observations (n = 50, i.e., 10 cases multiplied by five judges), a large effect (i.e., .80; cf. Cohen, 1988) can be detected with an acceptable probability (i.e., power of .80; cf. Moher, Dulberg, & Wells, 1994). However, due to the limited variability of data, such an average level of effect size corresponds to a rather small absolute difference: from 0.6 to 1.3 points on the 7-point Likert scale (i.e., accuracy: 0.85 points; efficacy of synthesis: 0.61 points; appropriateness: 1.04 points; clinical validity: 1.28 points).

Concerning the latter set of criteria (descriptive similarity for case description and interpretative similarity for case interpretation), a slightly more complex procedure was followed because of their nature; that is, we compared the ratings given by judges to the different pairs of clinical case analyses. More specifically, we grouped the 45 pairs into 3 types: pairs composed of two transcript-based clinical case analyses (TT pairs, n = 10); pairs composed of 2 ACASM-based clinical case analyses (AA pairs, n = 10); and mixed pairs, composed of one ACASM-based clinical case analysis and one transcript-based clinical case analysis (AT pairs, n = 25). (Figure 3 shows the composition of the types of pairs in terms of areas of the comparison matrix.) The three types of pairs were then compared.

The comparison of mixed pairs with pairs of transcript-based analysis (i.e., AT vs. TT) allowed to test if the average level of descriptive similarity and interpretative similarity was different between pairs of cases based on different data source formats (i.e., pairs belonging to the mixed AT cluster) and pairs of clinical case analyses based on the same data source format—where the event of the similarity should lead us to conclude that the data source formats (transcript and ACASM analysis) did not play a relevant role, so they can be used interchangeably. From a complementary point of view, the comparisons between AA pairs and TT pairs provides information on the ACASM’s capacity of making case analyses convergent with each other in terms of difference compared to the control group, namely, the level of convergence among transcript-based case study analyses. According to this standpoint, the AA vs. TT comparisons can be seen as a form of measurement of ACASM inter-rater reliability.
Since data regarding these criteria did not satisfy the conditions for parametric statistics, for all comparisons, non-parametric tests (Mann–Whitney U-test) were used. Moreover, in order to reduce the potential bias due to the small sample size, we chose to calculate exact p-values (Monte Carlo method) rather than asymptotic ones, and to adopt a less conservative level of $\alpha$ (.10) as the threshold of significance. Indeed, given the size of the sample ($n = 35$; i.e., 10 AA pairs vs. 25 AT pairs), such a level of $\alpha$ corresponds to a statistical power of about .80 of detecting a large effect size ($d = .80$). It has to be noted that such an effect size corresponds to absolute differences of 0.66 points and 0.76 points (on the 7-point Likert scale), for descriptive similarity and interpretative similarity, respectively.

**Results**

**Comparison Concerning Case-centered Criteria (Accuracy, Efficacy of Synthesis, Appropriateness, and Clinical Validity)**

Table IV shows the results of the random intercept model used for comparing ACASM-based and transcript-based analyses on the criteria entailing a case-centered evaluation (accuracy and efficacy of synthesis for case description; appropriateness and clinical validity for case interpretation).

For 3 out of 4 criteria—namely accuracy, efficacy of synthesis, and validity—fixed effects resulted in being significant, with transcript-based analyses higher than ACASM-based analyses. As to appropriateness, no significant differences were found between ACASM-based and transcript-based analyses (see Table IV).

**Comparison Concerning Pair-centered Criteria (Descriptive and Interpretative Similarity)**

The results are shown in Table V. Regarding the comparison between AT vs. TT pairs, these two groups did not differ—as expected—in interpretative similarity; on the contrary, TT pairs of clinical case analyses were found to have a significantly higher level of descriptive similarity than the AT pairs.

As to the comparison between AA vs. TT pairs, the hypothesis of a non-significant difference between the two groups could be confirmed for interpretative similarity, while AA pairs showed significantly higher levels of descriptive similarity than TT pairs.

**Discussion**

The results are partially consistent with the hypotheses. Indeed, differently from what we expected, transcript-based analyses showed better performances than
ACASM-based case analyses as to the level of case description. This resulted in all 3 criteria being adopted, that is: accuracy (i.e., the extent to which the case description is precise, clear, detailed, and systematic), efficacy of synthesis (i.e., the extent to which the case description optimizes the relationship between the quantity of the data presented and the quality of information produced), descriptive similarity (i.e., the extent to which the content of the two case descriptions overlapped)—and for both types of comparisons, that is: comparisons concerning case-centered criteria and comparisons concerning pair-centered criteria.

On the other hand, transcript-based and ACASM-based case analyses did not show differences at the level of case interpretation on 2 out of 3 criteria of comparison, that is: appropriateness (i.e., the extent to which the interpretation of the case was supported and justified by the case description) and interpretative similarity (i.e., the extent to which the content of the two case interpretations overlapped); differences were found for clinical validity (i.e., the extent to which the case interpretation is clinically meaningful, internally non-contradictory, and coherent from a clinical standpoint), as to the latter parameter, transcript-based analyses resulted higher than ACASM-based analyses.

In sum, results are not consistent with hypotheses as to the level of case description, while they are consistent, though only for two criterias out of three, with hypotheses as to the level of case interpretation. According to them, ACASM seems to provide a representation of the case (psychotherapy of Max) so that clinicians basing their work on it were able to elaborate clinical case interpretations—but not clinical descriptions—that are no different from those produced by clinicians who used verbatim transcripts directly. However, it has to be noted that this was true for two out of three criteria. Indeed, the transcript-based case interpretations showed a higher level of clinical validity than ACASM-based case interpretations.

Table IV. Random intercept model comparing ACASM-based and transcript-based analyses.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>ACASM-based estimatea</th>
<th>Std. Error</th>
<th>F</th>
<th>Sig</th>
<th>Random coefficient variationb</th>
<th>Std. Error</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>4.68</td>
<td>0.209921</td>
<td>792.998</td>
<td>.000</td>
<td>−1</td>
<td>0.296873</td>
<td>11.346</td>
<td>.001</td>
</tr>
<tr>
<td>Efficacy of synthesis</td>
<td>4.64</td>
<td>0.171075</td>
<td>1228.583</td>
<td>.000</td>
<td>−0.8</td>
<td>0.241937</td>
<td>10.934</td>
<td>.002</td>
</tr>
<tr>
<td>Validity</td>
<td>5.32</td>
<td>0.315278</td>
<td>444.467</td>
<td>.000</td>
<td>−1.24</td>
<td>0.445878</td>
<td>7.734</td>
<td>.008</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>5.4</td>
<td>0.292347</td>
<td>603.894</td>
<td>.000</td>
<td>−0.64</td>
<td>0.413441</td>
<td>2.369</td>
<td>.128</td>
</tr>
<tr>
<td>Amplitude</td>
<td>4.72</td>
<td>0.229536</td>
<td>782.107</td>
<td>.000</td>
<td>−0.52</td>
<td>0.318957</td>
<td>2.658</td>
<td>.110</td>
</tr>
</tbody>
</table>

aACASM-based estimate represents intercept values for each parameter.
bRandom coefficient variation represents the variation of transcript-based analysis as compared to ACASM-based estimates.

Table V. Comparisons between ACASM-based, transcript-based, and mixed pairs, concerning the pair-centered evaluation.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Type of pairs</th>
<th>Comparisons</th>
<th>M (SD)</th>
<th>M (SD)</th>
<th>M (SD)</th>
<th>Sig.a</th>
<th>Sig.a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pairs of ACASM-based clinical case analyses (AA) (n = 10)</td>
<td>Pairs of transcript-based clinical case analyses (TT) (n = 10)</td>
<td>Mixed pairs (AT) (n = 25)</td>
<td>Mixed pairs (AT) vs. pairs of transcript-based clinical case analyses (TT)</td>
<td>Pairs of ACASM-based analysis (AA) vs. pairs of transcript-based clinical case analyses (TT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Descriptive similarity</td>
<td>4.82 (.66)</td>
<td>4.10 (.42)</td>
<td>3.46 (.66)</td>
<td></td>
<td>.004</td>
<td>.011</td>
<td></td>
</tr>
<tr>
<td>Case interpretation</td>
<td>3.96 (.52)</td>
<td>3.22 (1.42)</td>
<td>3.33 (.81)</td>
<td></td>
<td>.957</td>
<td>.436</td>
<td></td>
</tr>
</tbody>
</table>

*Mann–Whitney U-test (one-tailed Monte Carlo exact test).
even though they seem to work in a different way as regards the kind of data they highlight for the interpretation.

In sum, partially consistently with our hypothesis concerning the level of case interpretation, ACASM proved to approximate the capability of verbatim transcripts to support the clinical understanding of the case.

On the other hand, though comparisons concerning the descriptive level of case analyses (accuracy, efficacy of synthesis, and descriptive similarity) are not consistent with the hypothesis, they lend themselves to be interpreted in a way that is not in contrast with the framework grounding the ACASM method. Indeed, the significant differences found at the descriptive level between ACASM-based and transcript-based case analyses can be explained as the consequence of the different formats of the two sources of data (i.e., transcripts and ACASM output).

Moreover, the fact that differences concern mainly the level of case description marks the fact that the different formats of the sources have a limited impact on the clinical interpretation of the case. This is particularly clear when results of the two pair-centered comparisons are considered: while TT pairs show a higher level of descriptive similarity than AT pairs, the same does not happen in the case of interpretative similarity (actually, as to the interpretative similarity, AT pairs resulted higher than TT, though not significantly). Incidentally, the latter result is an indication of the reliability of the judges, more specifically of their discriminative use of evaluation criteria.

A further result that is worth highlighting concerns comparisons between pairs composed of transcript-based clinical case analyses (TT pairs) and pairs composed of ACASM-based analyses (AA pairs). AA pairs were shown to be more similar than TT pairs in the case description. This result lends itself to be interpreted as a sign of the greater reliability of an ACASM-based case description. This conclusion is clearly consistent with the specificity of ACASM which, being a method based on automatized and invariant—though context-sensitive—algorithms (Salvatore et al., 2012), provides an output that, by definition, does not suffer from problems caused by the idiosyncratic ways human beings cannot but adopt when managing a large amount of data (like a verbatim transcript of a psychotherapy treatment). What the results at stake show is that ACASM works as a constraint on the clinicians’ case description, thus improving the reliability of data, as sustained by the level of inter-rater agreement (at least when it is esteemed in terms of similarity between descriptions).

**Conclusion**

The present study aimed at testing the capability of an automatized system of textual analysis of psychotherapy transcripts (ACASM) to provide clinicians with a case representation working as a valid input for clinical case analysis. To this end, two groups of clinicians were compared: a group grounding their clinical case analyses on the ACASM output and the other grounding their analyses directly on the transcripts.

Our results support—though not completely—the idea that ACASM’s synthetic and schematized representation of the thematic content of the psychotherapeutic exchange approximates the verbatim transcript as grounds for clinical interpretation of the case. This seems so despite the fact that the ACASM-based representation resulted different from verbatim transcripts as to the way of describing the cases.

**Limitations and Future Studies**

The following three main limitations have to be recognized. First, due to the fact that the transcript-based clinical case analysis is quite a time-consuming task, it was possible to involve only a rather limited number of clinicians in the study. This hampered the obtaining of a comparable distribution between the two groups as to the judges’ sex. Above all, as a consequence of the low numerosity, the statistical power of comparisons was quite low. Though we adopted statistical devices to reduce the impact of this aspect, it is not possible to fully exclude the possibility that our analyses have underestimated differences among transcript-based and ACASM-based clinical case analyses. However, given the low variability of data concerning most of the criteria the analyses did not prove to be fully lacking power; indeed, in two out of three comparisons concerning the interpretative level of analysis that produced non-significant results (i.e., appropriateness and interpretative similarity), the effect sizes they were able to detect corresponded to absolute differences that, in the final analysis, can be considered not so high (interpretative similarity: 0.76 points; appropriateness: 1.04 points, both on a seven-point Likert scale). Thus, even if these differences should be indicative of the better performance of transcript-based clinical case analyses, this could be seen as an acceptable “defeat” for ACASM: a minimal cost—in terms of less validity than the representation of the case provided by the verbatim transcripts—that is worth paying for the advantages of the method (i.e., less time- and energy-consuming, and higher reliability of analyses).
However, needless to say, we need further studies, endowed with more statistical power, in order to say something more definitive about the difference between ACASM and transcripts, in order to both estimate its magnitude better and understand the way of reducing it (e.g., by adopting a different format of ACASM output, in order to make it more comparable to the transcripts).

Second, the current study used one single case of psychodynamic psychotherapy as the data source. Future studies should aim at replicating results presented in this paper by analyzing psychotherapies of different orientations involving clients with different diagnoses.

Third, the clinical case analysis as it is defined in the present study (identification of recurrent themes and depiction of their temporal evolution [case description] and provision of a clinical interpretation of the case based on its description [case interpretation]) may be rather descriptive; future studies should try to verify whether ACASM may represent an adequate support for clinical case analyses of a higher complexity. This should provide the chance to better understand the magnitude and the meaning of the difference concerning the clinical validity of ACASM-based and transcript-based analyses as well as to find a way of reducing it.

### Implications for Clinical Research and Practice

This study has to be considered a further step on the route of the validation of ACASM as a bottom-up, context-sensitive automated method of textual analysis of psychotherapy transcripts, aimed at supporting the clinical interpretation of the case. In brief, while the previous study (Salvatore et al., 2012) showed that ACASM works in a way that is equivalent to the way human beings categorize the semantic content, this study showed that ACASM provides a representation of the case that grounds the clinical interpretation of the case in a way that approximates the one provided by the case’s verbatim transcript. Though affected by inherent limitations and not completely consistent with the hypothesis, such a finding has meaningful implications, which encourage us to consider this route worth pursuing, envisaging how ACASM lends itself to be developed in order to approximate even more the representation of cases provided by the verbatim transcript. Here what was written in the introduction to the first validation study of ACASM (Salvatore et al., 2012) is still valid:

> The development of bottom-up procedures of semantic analysis based on explicit, invariant rules of coding and yet able to take the contextuality of meaning into account. Procedures of this kind would represent a highly significant contribution to the growth of psychotherapy process research. On the one hand, they would allow the automated implementation of the semantic analyses. On the other hand, they would provide a shared ground supporting and constraining the (at least to date) non-renounceable human inferential judgments, so to increase the inter-coder agreement as well as the comparability among textual analysis. (p. 258)

It is worth adding that the automatized character of the method enables a dramatic increase of productivity, and thus an exponential enlargement of the empirical base of psychotherapy studies. Just to give an example, the verbatim transcript of Max’s case comprised a text of about 700 pages. We leave it to the reader to estimate the huge amount of time required just to read it. The application of ACASM to such a textual corpus required about 4 h, and this time is substantially independent from the length of the corpus. Thus, thanks to ACASM, the clinical researcher would be provided with a very large set of data, and so able to encompass many relevant clinical factors (concerning psychotherapists, patients, approaches, psychopathology, cultural contexts, and so forth) in a single analysis.

Finally, even if it has to be said that ACASM is essentially a device for clinical research, its validation can be useful for clinical practice and clinicians, too. One can imagine ACASM as a device to be used for monitoring the development of the thematic content of cases. Such monitoring could empower the handling of cases as well as being used in contexts of training, supervision, and self-learning. For instance, thanks to the ACASM output, a psychotherapist could check if there are similarities among her patients as regards the content and the evolution of the thematic nuclei, which would raise the question of the more or less implicit ways the therapist affects the client’s talk and its outcome. Again, the clinician can learn if certain thematic shifts work as markers of major changes within the clinical process. Something similar might be done in training contexts, for example as a supplementary tool to be integrated in supervision.

### Acknowledgments

We want to thank Diego Rocco, the therapist on Max’s case who put the verbatim transcripts at our disposal. We are grateful to Patrizia Calasso, Anna Dourdouma, Francesca Giordano, Rossano Grassi, Chiara Greco, Saba Khodayarifard, Elisa Latino, Antonella Mangia, Angela Mastrandrea, Giulia Miglietta, Silvia Potì, Antonio Rigliaco, and Cosimo Talò, who agreed to be involved in the
study as clinicians analyzing Max’s case or as judges rating the case analyses. Least but not the least, we want to thank the anonymous reviewer who suggested the use of the linear mixed-effect model as a way for increasing the statistical power of case-centered comparisons.

Notes

1 The complete set of materials is available upon request to the corresponding author of this paper.

2 We based the retrospective analysis of power (cf. Faul, Erdfelder, Lang, & Buchner, 2007 on Konstantopoulos (2009)’s procedure of estimation of the power of multilevel designs. Such a procedure uses Cohen’s (1988)’s power tables and computes the effect size parameter of the table as a function of the intra-class correlation at level-2, and the number of units within each level-2 unit.

References


