A Computer Program for Tracking the Evolution of a Psychotherapy Treatment

Bernard Maskit
Mathematics Department
Stony Brook University
Stony Brook NY 11794-3651
USA
daap@optonline.net

Wilma Bucci
Derner Institute
Adelphi University
Garden City NY 11530 USA
wbucci@optonline.net

Sean Murphy
New York Psychoanalytic Society
and Institute
247 East 82 Street
New York NY 10028, USA
smurphy1@gmail.com

Abstract

Describing troubling events and images and reflecting on their emotional meanings are central components of most psychotherapies. The computer system described here tracks the occurrence and intensity of narration or imagery within transcribed therapy sessions and over the course of treatments; it likewise tracks the extent to which language denoting appraisal and logical thought occurs. The Discourse Attributes Analysis Program (DAAP) is a computer text analysis system that uses several dictionaries, including the Weighted Referential Activity Dictionary (WRAD), designed to detect verbal communication of emotional images and events, and the Reflection Dictionary (REF), designed to detect verbal communication denoting cognitive appraisal, as well as other dictionaries. For each dictionary and each turn of speech, DAAP uses a moving weighted average of dictionary weights, together with a fold-over procedure, to produce a smooth density function that graphically illustrates the rise and fall of each underlying psychological variable. These density functions are then used to produce several new measures, including measures of the vividness of descriptions of images or events, and a measure of the extent to which descriptions of events or images and reflection on their meaning occur separately.

1 Introduction

In most forms of therapy, the treatment process includes two major phases of discourse: (1) the client talks about his or her concerns or problems and describes incidents related to these concerns, and (2) the client, perhaps with the help of the therapist, thinks about these concerns and evaluates the significance of the described incidents (Bucci, 2013). These phases are likely to be repeated in different contexts and with different contents. Some versions of psychodynamic therapy include reports of memories and dreams, as well as current interactions and interpretations of these. Some types of Cognitive-Behavioral Therapy include 'experiments' and other forms of 'homework' outside the treatment situation, and descriptions and evaluations of these critical events in the session. Some exposure therapies require that the client tell and retell the story of the trauma. Different treatments have different mixes of these two crucial phases.

These two styles of discourse have been termed *Symbolizing* and *Reorganizing* by Bucci (1997) and defined within the framework of her general theory of the referential process as this plays out in psychotherapy. According to this theory, an emotion schema is first aroused (this phase will not be discussed here); then communicated in the form of an image or narrative in the Symbolizing phase. The meaning of this image or story is then reflected on in the Reorganizing phase.

Much also occurs in a therapy session that lies outside these two modes of discourse; the client sometimes talks in general terms, sometimes is disfluent, and sometimes discusses matters outside the problem areas. In this paper, we describe a computer system designed to read texts, including transcriptions of therapy sessions, and track the extent

Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 134–145,
Denver, Colorado, June 5, 2015.

to which the speaker or writer is engaged in either of these two major phases or in some other mode.

The key components of our system are the Discourse Attributes Analysis Program (DAAP); the empirically derived Weighted Referential Activity Dictionary (WRAD) (Bucci and Maskit, 2006), which measures the extent to which the speaker or writer is in symbolizing mode; and the conceptually derived unweighted Reflection dictionary (REF), which measures the extent to which the speaker or writer is in reorganizing mode; in this paper we focus on these measures and measures derived from them. The system also includes other dictionaries, including disfluency and affect. These dictionaries can be used to help distinguish different phases of discourse (Kingsley, 2009), and also as measures of session effectiveness (Bucci and Maskit, 2007; Mariani et al., 2013; Andrei, 2011)[1].

According to Bucci (1997), Referential Activity (RA) is a psycholinguistic variable that concerns the extent to which language can capture a speaker's bodily, sensory and affective experience in such a way as to evoke corresponding experience in the listener. This communication generally takes the form of narratives or descriptions of imagery, and is the central indicator of the Symbolizing phase. The Weighted Referential Activity Dictionary (WRAD), which was designed to model RA, will be described in more detail below.

The Discourse Attributes Analysis Program (DAAP) is a modern text analysis program that produces, for each weighted or unweighted dictionary, and for each turn of speech or other user-defined segment a smoothly varying *density function* that tracks the rise and fall of the underlying psychological variable that the dictionary is designed to represent. DAAP uses the WRAD density function to derive a measure of average vividness while in symbolizing mode, and a measure of the extent of discourse spent in the symbolizing mode; DAAP also produces a measure of the extent to which a speaker's language is simultaneously in both symbolizing and reorganizing modes; there is evidence that a client's separation of these two modes of speech is related to session or treatment effectiveness.

[1]This system is publicly available for non-commercial use; it can be downloaded from ww.thereferentialprocess.org/the-discourse-attributes-analysis-program-daap.

There are several computer programs that have been used for the study of the content of psychotherapy sessions. Earlier programs, such as the General Inquirer (Stone et al. , 1966), as well as more recent comprehensive programs such as the LIWC of Pennebaker et al. (2001), use counts of words within user-defined segments, such as turns of speech, that match words in dictionaries defined by particular grammatical or psychological categories. Mergenthaler's Text Analysis System (1996) uses artificial segmentation into word blocks of approximately 150 words each, which enables some differentiation of different text modes. However this segmentation does not correspond to turns of speech or boundaries of meaning units. Some modern systems, as for example Salvatore et al. (2012), Imel et al. (2014) or Werbart et al. (2011) use topic models, Latent Semantic Analysis and/or other machine learning techniques to form their categories. Such programs are primarily concerned with the contents of discourse; some of these start by eliminating function words.

1.1 The Referential Process as a Common Mechanism in Talking Cures

Bucci (2013) argues that the sequence of Symbolizing and Reorganizing, characterized as the referential process, constitutes a common factor that occurs in different forms in a wide range of psychotherapies practiced today. In all these treatments, effectiveness of treatment depends on communicating emotional experiences in specific and concrete language. Such language has been shown by Bucci and Maskit (2007) to be associated with effective therapeutic work in psychodynamic treatment. In their extensive and critical review of process-outcome research, appearing in the current Handbook of Psychotherapy and Behavior Change, which provides the standard reference for the field of psychotherapy research, Crits-Cristoph, et al. (2013) have provided evidence that arousal of emotional experience, for example through retelling narratives of central traumatic events, is likely to be an essential ingredient in achieving positive outcomes in exposure treatments. They have also shown that concrete techniques, such as asking for specific examples of beliefs, also lead to better outcome in cognitive therapy, while abstract techniques were unrelated to subsequent improvement. Several studies reviewed by

Crits-Cristoph et al. (2013), have provided evidence that gains in self-understanding lead to improvements in symptoms in psychodynamic therapy. Reorganizing also occurs in the various forms of cognitive behavioral, schema and exposure treatments, in processes characterized as reappraisal, cognitive restructuring and development of techniques of self-regulation.

2 Dictionaries

The DAAP system uses several dictionaries to locate different phases of discourse. We are concerned here only with the Weighted Referential Activity Dictionary (WRAD), as a measure of the extent to which the speaker is in the Symbolizing phase, and the Reflection dictionary (REF), as a measure of the extent to which the speaker is in the Reorganizing phase.

2.1 Referential Activity

Variation in RA is interpreted as indicating a speaker's or writer's degree of emotional engagement or immersion in an experience as represented in language (Bucci, 1997). Such engagement is indicated by qualities of language ranging widely across divergent contents. A novelist may write about chasing the white whale, life in an English village in the early nineteenth century, or experiences in the Spanish Civil War with equivalent degrees of engagement; clients may describe a similarly wide range of experiences. The challenge in developing a lexical measure of engagement in experience was in capturing features of language that are dependent on style and essentially independent of content. Bucci and colleagues began development of the RA measure by turning to the principles of language style as given by Strunk and White, in particular their sixteenth principle of composition, which states: "Use definite, specific, concrete language." (1972) (pp. 16–18). Based on the features specified in this principle, four scales were developed: Specificity (quantity of detail), Clarity (organization and focus) Concreteness (degree of reference to sensory and other bodily experience) and Imagery (degree to which language evokes imagery). Definitions of the scales and procedures for rating them are outlined in a manual (Bucci et al. , 1992; Bucci and McKay, 2014). Scores for the four attributes are averaged to yield an overall RA measure for texts or text segments. The manual provides some explicit features of the several dimensions, but the scoring is based primarily on intuitive judgments. As for most linguistic processing, speakers of a language have more implicit knowledge concerning language style and its effects than they are able to state in explicit terms. Scorers achieve acceptable reliability levels by reading the manual and brief training with practice segments.

The RA scales have been applied to many types of texts, including brief monologues, early memories, and Thematic Apperception Test (TAT) protocols as well as transcripts of therapy sessions, in populations varying on demographic and clinical dimensions. In a meta-analysis of 23 studies, Samstag (1996) found significant relationships, with moderate to strong effect size, between RA scales and other indicators of capacity to connect cognitive, and emotional experience to language. While the scales are reliably scored with relatively brief training, computerized procedures are needed to enable assessment of RA in large sample and longitudinal studies, and micro-analytic tracking of fluctuation in RA within various forms of communicative discourse. Traditional methods of computerized language analysis depend on construction of word lists representing specified contents and concepts. For the RA dimension, a different approach to modeling the scales was used.

A first computer model of Referential Activity, called CRA, was empirically derived by Mergenthaler and Bucci (1999) using a set of transcriptions of spoken language that had been scored for RA. The model consisted of two dictionaries; one made up of words that are used significantly more often in high RA speech and the other of words used significantly more often in low RA speech. These were used as a measure of RA with the Text Analysis System (TAS) of Mergenthaler (1996), which segments each text into word blocks of approximately 150 words each, and then computes a mean CRA score for each such word block (High RA words minus Low RA words divided by the total number of words).

The dictionary currently in use, the Weighted Referential Activity Dictionary (WRAD), was also empirically derived from a set of transcriptions of spoken language that had been scored for RA (Bucci

and Maskit, 2006). Weighted referential activity dictionaries have also been constructed in Spanish (Roussos and O'Connell, 2005), and Italian (Mariani et al., 2013). The WRAD contains approximately 700 single-word items, including many very frequent function words; thus WRAD covers roughly 75–85% of spoken language. Each item in the WRAD has a weight, ranging from -1 to $+1$, that was empirically derived so as to model that item's usage in segments at different RA levels. For example, an item with weight -1 is used much more often in text segments having RA scores in the range of 0 to 2.75, an item with weight $+1$ is used much more often in text segments having RA scores in the range of 7.25 to 10. As described in Bucci and Maskit (2006), the algorithm used to make the WRAD uses different definitions of the term 'much more often' to construct different dictionaries; the final one is chosen by maximizing the correlation with judge's scores of RA on a separate set of texts.

As shown in Bucci and Maskit (2006), the WRAD and CRA were tested on a set of 113 text segments that had been scored for RA, and that had not been used in the construction of either dictionary. For this test set, the WRAD/RA correlation was .47; the CRA/RA correlation was .31. As the coverage of a dictionary could be important for interpreting the corresponding density function, we note that the CRA coverage of this material was .50, while the WRAD coverage was .83. To the best of the authors' knowledge, there are no other computer measures of RA to test the WRAD against.

Since the contents of the WRAD are based on the scales, which are intuitively scored, the weights are generally independent of linguistic or grammatical category and relate to language style in ways that are generally not explicitly understood. Thus general content or grammatical categories, such as are applied in the LIWC and other text analysis systems, could not be used in making the WRAD. For example, 'in' and 'inside' might be grouped together in typical categorical systems; however the WRAD weight of the word 'in' is $+1$, signifying that people generally use this word far more often in symbolizing mode than otherwise; the weight of the word 'inside' is -1, signifying that people generally use this word far less often in symbolizing mode than otherwise. Similarly, the words 'and', 'was', 'she' and 'on' each have the highest possible WRAD weight of $+1$, while the words 'also', 'is', 'it' and 'off', which appear semantically related to these four items respectively, have very low WRAD weights ('it' has weight $-.875$, the others have the lowest possible weight of -1). The content words in the dictionary include 'mother' and 'class', which have WRAD weight $+1$, as well as 'family' and 'money', which have WRAD weight -1.

Post-hoc examination of the lexical contents of the WRAD suggests that many frequent words with high WRAD weights are those with the types of functions required for telling stories. The five most frequent words with weights of +1 are the conjunction 'and', the definite article 'the', the past tense verb 'was', the spatial preposition 'in', and the personal pronoun 'she'; these are items with the types of pointing and connecting functions that are likely to be used in describing episodes — to locate the objects of discourse in place and time, and to join together or relate objects or ideas — as well as past tense verbs that serve as indicators of memory retrieval, and third person singular animate pronouns that are used to refer to specific other people figuring in an episode. The most frequent words with low WRAD weights are associated with subjective focus ('I') rather than pointing to objects and describing events, present rather than past tense ('is'), general and abstract usage ('it' and 'that') and disfluency indicated by the filled pause term ('mm') (Bucci et al., 2015). Other factors contributing to the contents of the WRAD are now being studied by Murphy et al. (2015).

2.2 Reflection

The Reflection dictionary (REF) is an unweighted list of over 1400 words that relate to reflection or logical thought. These include logic terms ('if', 'but'); words referring to cognitive functions ('think', 'plan'), or entities ('cause', 'belief'); problems of failure of cognitive or logical functions ('confuse', 'confound'); complex verbal communicative functions ('comment', 'argue'); and features of mental functioning ('creative', 'logical').

The REF dictionary was formed by having three judges, using a definition of the Reflection category, independently rate words from a large set of texts, including the texts used to make and test the WRAD.

For each word, if all three judges agreed on its inclusion, it was added to the REF dictionary. If two of the three agreed on inclusion, the word was given to a fourth judge, and included in the dictionary if the fourth judge agreed.

3 The Discourse Attributes Analysis Program (DAAP)

The DAAP system operates on the assumption that each dictionary represents an underlying psychological process that varies over time. For each dictionary and each turn of speech DAAP produces a smoothly varying *density function* that models the underlying psychological variable. DAAP uses these density functions to produce several derived measures; the density functions and some derived measures will be described and illustrated below.

The WRAD weights are given as lying between -1 and $+1$, with a *neutral value* of 0, corresponding to the RA scale score neutral value of 5. As is usual for a text analysis system, DAAP assigns the weight 0 to a word that is not in a dictionary; a word that is in an unweighted dictionary is assigned the weight $+1$; a word that is in a weighted dictionary is assigned its dictionary weight. As negative values are sometimes difficult to interpret for psychological variables, the WRAD dictionary scores are linearly transformed so as to lie between 0 and 1, with neutral value at $.5$. With this transformation, the DAAP density functions are all non-negative and have values between 0 and $+1$.

In what follows, the WRAD neutral value of .5 is used as a dividing line between segments of discourse that are considered to be high in RA and those that are considered to be low. This division enables DAAP to segment text into contiguous sets of words for which the WRAD density function is either high or low; that is, greater than or less than this neutral value.

3.1 Ordinary Text Analysis Functions

Session material is usually transcribed with markers indicating changes in speaker. DAAP permits but does not require this or other segmentation markers and treats each such marker as indicating a new turn of speech, thus allowing for different definitions of 'turn of speech'. For example, pauses of a certain length or longer might be viewed as indicating a new turn of speech, even if no change of speaker has actually occurred; or certain interjections, such as 'umhm', might be viewed as not indicating a change of speaker. For qualitative analysis of content spoken at interesting points as indicated by the graphs of the density functions, DAAP produces a marked text; this reproduces the original text file with markers inserted every 50 words.

3.2 The Density Function

For each dictionary and each turn of speech, DAAP constructs a *density function*, which has a non-negative value at each word in the turn of speech. This construction starts with a moving weighted average of the dictionary weights, where the weighting function is an exponential closely related to the normal curve. This weighting function is equal to zero for all values outside the range, $-99 \leq x \leq +99$. Except for the first and last 99 words of each turn of speech, the density function is equal to this moving weighted average. Special adjustments using a fold-over technique are made for the first and last 99 words. These adjustments have the consequence that the mean of the density function is equal to the mean of the dictionary weights. Precise definitions of the density function and the measures outlined below are given in the appendix.

Most therapy sessions have a total of between 5,000 and 7,000 words. For each dictionary, the density function appears as a visually smooth curve with discontinuities at each change of speaker. (As explained in the appendix, one can regard the density function as being defined at every real number so that it is a mathematically smooth function for every turn of speech.) The segments where the WRAD density function lies above the neutral value of .5 are easily located, and the text corresponding to these segments can be located in the marked text.

We illustrate the density function and the derived measures with graphs of the WRAD and REF density functions of Session 4 from a treatment carried out by Carl Rogers at the research program in psychotherapy of the University of Chicago; the client is known as Miss Vib (Rogers and Kinget, 1965). The treatment was regarded as highly successful, and this session was considered a pivotal session. Rogers and Kinget say that during Session 4 "the inner disorga-

nization that characterizes this phase of the process reaches its climax" leading then to a shift into an evaluation mode in Session 5. In both these figures, the client data appears as the thinner black line; the therapist data appears as the thicker black line.

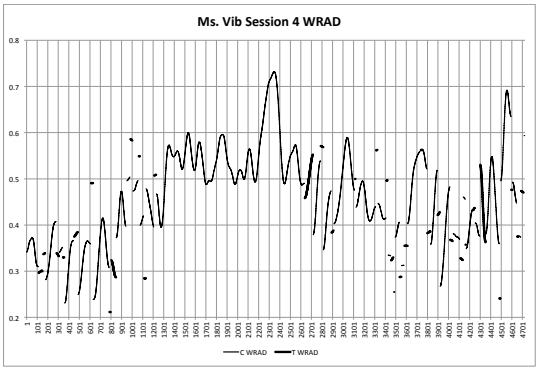

Figure 1: Client and therapist WRAD density.

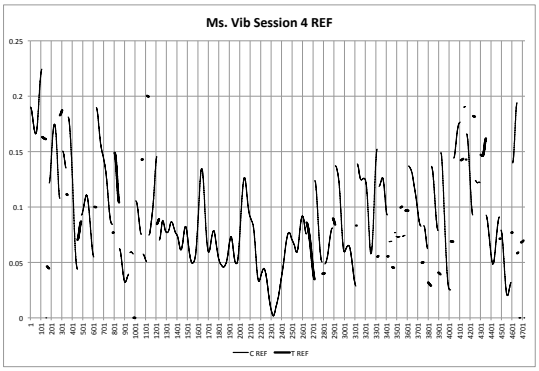

Figure 2: Client and therapist REF density.

The session opens with the client expressing her concern that she is not accomplishing much in the treatment, she wonders whether she should be doing something different, she's not sure what she should talk about. She feels she was functioning well several years ago; she doesn't know what is blocking her now in her life. The conversational pattern of relatively brief utterances by the client and responses by the therapist is characteristic of the treatment and continues until about word 1200; the therapist reflects the client's ideas, with some shifts in language. Following these interactions, Miss Vib begins a turn of speech of approximately 1400 words, lasting from about word 1200 to about word 2600; this is the longest uninterrupted utterance in this treatment.

In this segment, which reaches a WRAD peak of .73, as shown in Figure 1, the client tells how she had accepted a fellowship for graduate training without realizing what the fellowship required; then tells a detailed and vivid story of how she had to push herself through a project that she did not want to do, that she did not believe in, and that required work with a population and in a setting that was frightening for her. She is deeply immersed in the description and it is highly evocative for the reader.

During the same period, the Reflection (REF) measure is very low, as shown in Figure 2. As WRAD declines following the extended speech segment, REF increases; the graphs of the WRAD and REF density functions are close to mirror images of one another. This configuration, indicating separation of the Symbolizing and Reorganizing phases, is a major marker of the referential process. Miss Vib tells a pivotal story, and then reflects on it, leading to development of new emotional meanings.

3.3 Derived Functions

The *covariation* between two variables is a measure of the degree to which the variables are simultaneously high and low. Mathematically it is exactly the same as the (Pearson) correlation coefficient between the corresponding density functions. As the values of a density function at nearby words are not statistically independent, we call this operation *covariation* rather than correlation. The covariation of REF and WRAD is an indicator of the extent to which the speaker is separating the functions of symbolizing and reorganizing; we expect this measure to be mainly negative and to be more negative in more effective sessions and treatments (see Sec. 4.2). The REF-WRAD covariation for the 1405 words in the client's extended turn of speech shown above is -.76; for the session as a whole the covariation is -.56.

The *High WRAD Proportion (HWP)* is computed for each turn of speech, or for any user-defined set of turns of speech, as the proportion of words for which

the WRAD density function lies above its neutral value of .5. It is used as an indicator of the proportion of time in a session that the client is in symbolizing mode. We expect this measure to be high for client speech in effective sessions, and at least not to decrease over time in successful treatments (see Sec. 4.2).

The *Mean High WRAD (MHW)* is the mean of the difference between the WRAD density function and the WRAD neutral value of .5, when this difference is positive. That is, MHW is computed by considering only those words for which the WRAD density function is greater than its neutral value of .5. It is used as an indicator of the intensity or vividness of language when the speaker is in symbolizing mode, and is independent of the number of words in the turn of speech or other text segment(s) under consideration. As with HWP, we expect this measure for client speech to be relatively high in more effective sessions and to be at a generally high level in successful treatments (see Sec. 4.2).

The figures above illustrate the power of the density functions to identify pivotal moments of a session. For the long turn of speech discussed above, Mean WRAD (MWRAD) = .55, HWP = .79, MHW = .07 and Mean REF = .07. For this session as a whole, the client Mean WRAD = .47, HWP = .40, MHW = .07 and Mean REF = .09.

4 Related Research

4.1 Evidence for Construct Validity

A relationship between WRAD and narrativity was established by Nelson et al. (2008), who used a set of 55 narratives from high school students talking about their most stressful time. They found a high (Spearman) correlation ($\rho = .69$, $p < .01$) between Mean WRAD and a measure of narrativity given by a count of temporal sequences (Labov, 1997).

Using a data set provided by Addis et al. (2008), Maskit et al. (2015) found a relationship for both MWRAD and HWP with a measure of episodic memory given by the proportion of 'internal' to total 'details'; the measures were applied to a set of responses by 32 participants to prompts for 8 past and 8 future personal (episodic) events. The responses were recorded, transcribed and separated into details by Addis et al. (2008). A detail was considered to be internal if it was a specific fact concerning the main event being described, and was considered to be external if it was general rather than specific, or it concerned an event other than the main event or was a repetition. For the 32 subjects, high (Pearson) correlations were found between this measure of episodic memory and HWP ($r = .68$, $p < .01$) and with MWRAD ($r = .58$, $p < .01$).

A set of 70 segments taken from psychoanalytic sessions were rated by judges on a scale of 1 to 7 for location in each of the phases: Arousal, Symbolizing and Reorganizing. For the symbolizing phase, high (Pearson) correlations were found between those ratings and MWRAD ($r = .56$, $p < .01$), HWP ($r = .58$, $p < .01$) and MHW ($r = .55$, $p < .01$); a high negative correlation with REF ($r = -.27$, $p = .02$) was also found. For the reorganizing phase, a high positive correlation with REF ($r = .42$, $p < .01$), and high negative correlations with MWRAD ($r = -.60$, $p < .01$), HWP ($r = -.60, p < .01$) and MHW ($r = -.52, p < .01$) were also found (Kingsley, 2009).

Murphy (2015) presents three studies showing that WRAD scores tend, on average, to be substantially higher when participants are asked to discuss stories, events, or other scenarios such as dreams in comparison to other speech contexts ($1.5 \leq d \leq 3.5$). These studies also show that WRAD scores have moderate temporal stability over a six week period for the same task ($.33 \leq r \leq .61$).

4.2 Applications to Psychotherapy

In a study of 16 sessions from a long term psychoanalysis, Bucci and Maskit (2007) found high (Pearson) correlations between a measure of session effectiveness based on clinical judgments (Freedman et al., 2003) and DAAP measures; these include the negative REF-WRAD covariation ($r = .70$, $p < .01$), and MWRAD ($r = .54$, $p < .05$). These suggest that in the more effective sessions, the client had more separation of symbolizing and reorganizing discourse, and was more vivid while in symbolizing mode.

Using the Italian version of this system, Mariani et al. (2013) used Spearman correlations to examine the client speech for entire sessions of three successful psychotherapies. They found as expected that HWP increased over time; that is, the client spent an

increasing proportion of time in symbolizing mode [($N = 10$, $\rho = .79$, $p < .01$), ($N = 33$, $\rho = .43$, $p < .01$), ($N = 23$, $\rho = .33$, $p = .07$)]; the overall Mean WRAD (MWRAD) also increased over time for all three treatments, [($N = 10$, $\rho = .60$, $p < .05$), ($N = 33$, $\rho = .50$, $p < .01$), ($N = 23$, $\rho = .36$, $p < .05$)], and the REF-WRAD covariation decreased over time; that is, the client on average had more separation of the functions of symbolizing and reorganizing, [($N = 10$, $\rho = -.48$, $p = .08$), ($N = 33$, $\rho = -.49$, $p < .01$), ($N = 23$, $\rho = -.44$, $p < .05$)].

Andrei (2011) studied 15 sessions of a successful psychotherapy treatment (as measured by standard client self-report measures). Using Spearman correlations for client speech only and for sessions as a whole, she found predicted increases in MHW ($\rho = .52$, $p < .05$); MWRAD ($\rho = .37$) and the HWP ($\rho = .35$).

In a study of 14 sets of candidate treatment notes from the New York Psychoanalytic Society and Institute Treatment Center, high (Pearson) correlations were found between a measure of treatment effectiveness (found by comparisons of client functioning between beginning and end of treatment) and both MHW ($r = .73$, $p < .01$) and MWRAD ($r = .70$, $p < .01$), for the treatment notes as a whole (Bucci et al. , 2012).

5 Limitations and Future Research

The system presented here for the study of psychotherapy process is based on Bucci's theory of the referential process (1997). We are concerned with measurements for two of the three phases of this process, Symbolizing and Reorganizing. The WRAD, which was designed as a measure of the Symbolizing phase has been extensively validated and has been favorably compared with the only other known measure of this psycholinguistic style, the CRA. WRAD's correlation to the scales might be improved by including some number of less frequently used words, as was done with the Italian WRAD (Mariani et al., 2013), and/or by enlarging the number of text segments scored for RA on which the measure is based and using some machine learning techniques.

The Reflection dictionary, used to mark the Reorganizing phase, is unweighted and theoretically based. Our current information concerning the REF-WRAD covariation suggests that, just as people use different function words to different extents when speaking at different levels of the Symbolizing phase, so they may also use different function words to different extents for different aspects of the Reorganizing phase. To the best of the authors' knowledge, no weighted reorganizing dictionary or set of dictionaries based on these ideas has as yet been developed.

We have not here addressed the Arousal phase of the referential process, in part due to limitations of space, and in part because this phase is sometimes marked by silence, variation in speech rate and acoustic features rather than lexical items. The system described here, based on word count, includes a Disfluency measure, which can to some extent be used to mark the Arousal phase. We are currently developing a Variable Time DAAP (VTDAAP) that uses sound recordings to provide acoustic data, such as changes in pitch and intensity as well as pausing and speech rate. VTDAAP produces data for which the independent variable is time rather than word count. A first version of this program has been tested and is currently being revised; we expect it to be publicly available in early 2016.

A major feature of the DAAP system is the production of density functions. These depend on the values of the parameters used for the weighting function, as described in the appendix. These parameters were chosen so as to make the graphs of the WRAD and REF density functions for psychotherapy sessions reasonably smooth and readable. Changes in these parameters would produce changes in the derived functions described above; as there are, however, no other measures of the variables these measures are meant to model, we have no standard against which to measure the effect of changing the weighting function parameters.

Several new studies are currently under way relating WRAD to narrativity and Episodic Memory; these use a new version of DAAP that produces density functions based on user-defined segmentation.[2]

[2]We expect this version of DAAP to be publicly available in 2016.

References

Donna Rose Addis, Alana T. Wong and Daniel L. Schacter. 2008. Age-Related Changes in the Episodic Simulation of Future Events. *Psychological Science*,19(1):33–41.

Claudia G. Andrei. 2011. A multi-perspective analysis of a psychodynamic treatment. Doctoral dissertation retrieved from ProQuest (AAT 3491569).

David Arvidsson, Sverker Sikström and Andrzej Werbart. 2011. Changes in self and object representations following psychotherapy measured by a theory-free, computational, semantic space method. *Psychotherapy Research* 21(4):430–446.

Wilma Bucci. 1997. *Psychoanalysis and Cognitive Science: A multiple code theory.* Guilford Press, New York NY

Wilma Bucci. 2002. The Referential Process, Consciousness, and the Sense of Self. *Psychoanalytic Inquiry*, 22:766–793.

Wilma Bucci. 2013. The referential process as a common factor across treatment modalities. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 16:16–23.

Wilma Bucci, Rachel Kabasakalian and the RA Research Group. 1992. *Instructions for scoring Referential Activity (RA) in transcripts of spoken narrative texts.* Ulm, Germany, Ulmer Textbank.

Wilma Bucci, Rachel Kabasakalian and the RA Research Group. 2004. *Instructions for scoring Referential Activity (RA) in transcripts of spoken narrative texts.*

Wilma Bucci and Bernard Maskit. 2006. A weighted dictionary for Referential Activity. In James. G. Shanahan, Yan Qu and Janyce Wiebe (Eds.) *Computing Attitude and Affect in Text*. Springer, Dordrecht, The Netherlands; 49–60.

Wilma Bucci and Bernard Maskit. 2007. Beneath the surface of the therapeutic interaction; The psychoanalytic method in modern dress. *Journal American Psychoanalytic Assn.*, 55, 1355–1397. doi:10.1177/000306510705500412.

Wilma Bucci, Bernard Maskit and Leon Hoffman. 2012. Objective Measures of Subjective Experience; The use of Therapist Notes in Process-Outcome Research. *Psychodynamic Psychiatry*, 40(2):303–340.

Wilma Bucci, Bernard Maskit and Sean Murphy. 2015. Connecting emotions and words: the referential process. *Phenomonology and the Cognitive Sciences*, doi:10.1007/s11097-015-9417-z.

Wilma Bucci and Rachel K. McKay. 2014. Manual for Scoring RA Scales. figshare http://dx.doi.org/10.6084/m9.figshare.962956.

Paul Crits-Christoph, Mary Beth C. G. Gibbons and Dahlia Mukherjee. 2013. Psychotherapy Process-Outcome Research. In Michael J. Lambert. *Bergin and Garfields's Handbook of Psychotherapy and Behavior Change*, Sixth Edition. (pp. 298–340) Hoboken NJ: John Wiley and Sons, Inc.

General Inquirer. www.wjh.harvard.edu.

Norbert Freedman, Richard Lasky and Marvin Hurvich. 2003. Two pathways torwards knowing psychoanalytic process. In M. Leuzinger-Bohleber, A.U. Dreher and J. Canestri (eds.) *Pluralism and unity: Methods of reserach in psychoanalysis*, International Psychoanalytical Association, pp. 207–221.

Zac E. Imel, Mark Steyvers and David C. Atkins. 2014. Computational Psychotherapy Research: Scaling up the Evaluation of Patient-Provider Interactions. *Psychotherapy*, Advance online publication. http://dx.doi.org/10.1037/a0036841.

George Kingsley. 2009. The clinical validation of measures of the Referential Process. Retrieved from Dissertations & Theses Adelphi University. (Publication No. AAT 3377938)

W. Labov. 1997. Some further steps in narrative analysis. *Journal of Narrtive and Life History*,7: 395–415.

Rachele Mariani, Bernard Maskit, Wilma Bucci and Alessandra DeCoro. 2013. Linguistic measures of the referential process in psychodynamic treatment: The English and Italian versions. *Psychotherapy Research.*, 23(4):430–447; doi:10.1080/10503307.2013.794399

Bernard Maskit. 2014. The Discourse Attributes Analysis Program (DAAP) Operating Instructions. figshare. http://dx.doi.org/10.6084/m9.figshare.947740.

Bernard Maskit, Wilma Bucci and Sean Murphy. 2015 Computer Based Measures of Referential Activity and their Use as Measures of Episodic Memory. www.thereferentialprocess.org/theory/episodic-memory-and-referential-activity.

Erhard Mergenthaler. 1996. Emotion-Abstraction patterns in Verbatim Protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 64:1306–1315.

Erhard Mergenthaler and Wilma Bucci. 1999. Linking verbal and nonverbal representations: Computer analysis of Referential Activity. *British J. of Medical Psychology*, 72:339–354.

Sean Murphy. 2015. Measuring the Details that Matter: The Validity and Psychometric Properties of the Weighted Referential Activity Dictionary. (Under Review)

Sean Murphy, Wilma Bucci and Bernard Maskit. 2015. Putting Feelings into Words: Cross-Linguistic Markers of the Referential Process. This Volume.

Kristin L. Nelson, Damian J. Moskovitz and Hans Steiner. 2008. Narration and Vividness as Measures of Event-Specificity in Autobiographical Memory. *Discourse Processes*, 45:195–209.

J. W. Pennebaker, M. E. Francis and R. J. Booth. 2001. *Linguistic Inquiry and Word Count (LIWC): LIWC2001.* Mahwah, NJ, Lawrence Erlbaum Associates.

Carl R. Rogers and M. G. Kinget. 1965. *Psychothérapie et relations humaines. Théorie et pratique de la thérapie non-directive.* Ed. Nauwelairts Louvain

Andres J. Roussos and M. O'Connell. Construcción de un diccionario ponderado en español para medir la Actividad Referencial. *Revista del Instituto de Investigaciones de la Facultad de Psicologa. UBA*, 10(2):99–119.

Sergio Salvatore, Alessandro Gennaro, Andrea F. Auletta, Marco Tonti and Mariangella Nitti. 2012. Automated method of content analysis: A device for psychotherapy process research *Psychotherapy Research*, 22(3):256–273.

Nicholas Samstag. 1996. A meta-analysis of referential activity. (Doctoral Dissertation) Retrieved from ProQuest (9705478).

P. J. Stone, D. C. Dunphy, M. S. Smith and D. M. Ogilvie. 1966. *The general inquirer: A Computer approach to content analysis.* Cambridge MA and London, The MIT Press.

W. Strunk Jr. and E. B. White. 1972. *The elements of style.* The Macmillan Company.

6 Appendix: The Density Function and Associated Measures

In this appendix we give precise definitions of the density function and the DAAP measures derived from it. The density function is constructed in two steps; the first is a moving weighted average of the dictionary weights using an exponential function related to the normal curve as the weighting function. It is a general statement that the mean of a moving weighted average is equal to the mean of the original function; however, the moving weighted average may have non-zero values at points where the original function is not defined. That is, if we are looking at dictionary values defined for the points, $0, \ldots, N$, and the weighting function is different from 0 for the points $-m < x < +m$, then the moving weighted average may have non-zero values at the points $-m + 1 \leq x \leq N + m - 1$. The second step in the construction of the density function is a *fold-over* procedure that adjusts the values of the moving weighted average to take these 'extra' values into account. After the second step is accomplished, the density function is defined exactly for the words $0, \ldots, N$, and its mean is equal to the mean of the original function.

The *Weighting Function* for the moving weighted average is defined in terms of two parameters: a 'pointedness' parameter q, and a 'support' parameter m. We start with the function W_0, which is zero for all x outside the range $-m < x < +m$. Inside this range

$$W_0(x) = \exp(-qm^2 \frac{m^2 + x^2}{(m^2 - x^2)^2}). \qquad (1)$$

Let $S = \sum_{x=-m+1}^{m-1} W_0(x)$. Then the weighting function $W(x) = W_0(x)/S$.

This weighting function has the following properties:

- $\sum W(x) = 1$.
- W is centered at 0, where it attains its maximum; its graph is symmetric with respect to the y-axis. (It is an even function.)
- W is strictly increasing from $-m$ to 0 and strictly decreasing from 0 to m.

- W is equal to zero for all x outside the range $-m < x < +m$.

For most purposes used here, the values of q and m are taken to be $q = 2$ and $m = 100$. We assume below that these are the values of q and m.

Using the same formula, we could have defined W_0 for all real numbers x, where it is positive exactly in the range $-m < x < m$, and then replaced the sum S by the corresponding integral. In this way W would be defined for all x, would have the properties listed above, and would also have derivatives of all orders at all points.

Let R denote some dictionary function defined on the set of words numbered from 0 to N, where $R(x) \geq 0$ for all x in this range. We also set $R(x) = 0$ for x outside this range. The first approximation to the R density function is the moving weighted average (convolution product)

$$C_R(x) = \sum_{y=-m+1}^{m-1} R(y-x)W(y). \qquad (2)$$

This is a finite sum for every x, and is equal to zero for all x outside the range $-100 < x < N + 100$. As remarked above, the mean of C_R, taken inside this range, is equal to the mean of R in the range $0 \leq x \leq N$. As it is difficult to assign a meaning to the value of C_R outside the range $0 \leq x \leq N$, some adjustments are needed to account for these values. The adjustments described below are equivalent to the idea that we "fold over" the x-axis, along with the graph of C_R at the points $x = -1/2$ and $x = N + 1/2$; then add these folded over values to the original values of C_R; and repeat this process as often as necessary.

As a particular example, assume that $N = 300$. Then C_R is defined and non-negative for the points, $-99 \leq x \leq 399$. For this example, we can write down the density function D_R as follows:

$$\begin{aligned}
D_R(0) &= C_R(0) + C_R(-1),\\
D_R(1) &= C_R(1) + C_R(-2), \ldots,\\
D_R(100) &= C_R(100), \ldots,\\
D_R(200) &= C_R(200), \ldots,\\
D_R(299) &= C_R(299) + C_R(302),\\
D_R(300) &= C_R(300) + C_R(301).
\end{aligned}$$

To make this process precise, we first define the auxiliary function $\widetilde{C}(R)$, by introducing the reflections: $r_1(x) = -x - 1$, which is reflection about the point $x = -1/2$, and $r_2(x) = -x+2N+1$, which is reflection about the point $x = N+1/2$. We consider the group of motions G of the real line generated by r_1 and r_2, and define the auxiliary function $\widetilde{C}_R$ by

$$\widetilde{C}_R(x) = \sum_{g \in G} C_R(g(x)). \qquad (3)$$

This is a finite sum for every integer x. This function $\widetilde{C}_R$ is invariant under the group G; that is, for every x and for every $g \in G$, $\widetilde{C}_R(x) = \widetilde{C}_R(g(x))$.

Finally, the density function D_R is defined by $D_R(x) = \widetilde{C}_R(x)$ for $0 \leq x \leq N$, and $D_R(x) = 0$ for x outside this range. With this definition, the mean of the density function D_R is equal to the mean of the dictionary values, R; that is $\sum_{x=0}^{N} D_R(x) = \sum_{x=0}^{N} R(x)$.

Let D_W be the density function for WRAD; once this density function has been defined, it is easy to describe the High WRAD Proportion (HWP) and the Mean High WRAD (MHW).

Let V be the set of all integers x in the range $0 \leq x \leq N$ for which $D_W(x) > .5$, and let Z be the number of points in V; so that $0 \leq Z \leq N + 1$. Then

$$HWP = Z/(N + 1), \qquad (4)$$

and

$$MHW = \sum_{x \in V} (D(x) - .5)/Z. \qquad (5)$$

The covariation $C(D_1, D_2)$ between two distinct density functions, D_1 and D_2, both defined on the same set of words labeled $0, \ldots, N$, is then defined exactly as the Pearson correlation coefficient (provided both densities are not constant):

$$C(D_1, D_2) = \frac{\sum_{x=0}^{N}(D_1(x) - M_1)(D_2(x) - M_2)}{\sqrt{V_1 V_2)}},$$

where M_1, respectively M_2, is the mean of D_1, respectively D_2, and V_1, respectively V_2, is the variance of D_1, respectively D_2.

The graph of the density function D for each turn of speech appears as a smooth curve. This can be explained by the underlying mathematical theory, which uses the continuous version of the weighting function W. Here, the dictionary values $R(x)$ are extended so as to be defined for all x in the

range $-1/2 \leq x < N + 1/2$, by requiring that, for each integer n, where we already have $R(n)$ defined, we set $R(x) = R(n)$ for all x in the interval $n - 1/2 \leq x < n + 1/2$. Then the moving weighted average $C_R(x)$ is defined as in equation 2, replacing the sum by an integral. The definitions of $\widetilde{C}_R$ and D_R then follow exactly as above. One can use this continuous definition of the density function to define MHW, HWP and the covariations by appropriate modifications of the above formulae; that is, by replacing sums with integrals, and by replacing counts of words by sums of lengths of intervals.