Published in final edited form as:

J Fam Psychol. 2012 October ; 26(5): 816–827. doi:10.1037/a0029607.

Topic models: A novel method for modeling couple and family text data

David C. Atkins,

Department of Psychiatry and Behavioral Sciences, University of Washington

Tim N. Rubin,

Department of Cognitive Science, University of California, Irvine

Mark Steyvers,

Department of Cognitive Science, University of California, Irvine

Michelle A. Doeden,

Department of Clinical Psychology, Fuller Theological Seminary

Brian R. Baucom, and

Department of Psychology, University of Southern California

Andrew Christensen

Department of Psychology, University of California, Los Angeles

Abstract

Couple and family researchers often collect open-ended linguistic data – either through free response questionnaire items or transcripts of interviews or therapy sessions. Because participant's responses are not forced into a set number of categories, text-based data can be very rich and revealing of psychological processes. At the same time it is highly unstructured and challenging to analyze. Within family psychology analyzing text data typically means applying a coding system, which can quantify text data but also has several limitations, including the time needed for coding, difficulties with inter-rater reliability, and defining a priori what should be coded. The current article presents an alternative method for analyzing text data called topic models (Steyvers & Griffiths, 2006), which has not yet been applied within couple and family psychology. Topic models have similarities with factor analysis and cluster analysis in that topic models identify underlying clusters of words with semantic similarities (i.e., the “topics”). In the present article, a non-technical introduction to topic models is provided, highlighting how these models can be used for text exploration and indexing (e.g., quickly locating text passages that share semantic meaning) and how output from topic models can be used to predict behavioral codes or other types of outcomes. Throughout the article a collection of transcripts from a large couple therapy trial (Christensen et al., 2004) is used as example data to highlight potential applications. Practical resources for learning more about topic models and how to apply them are discussed.

Keywords

Text; Linguistic Analyses; Topic Models; Couple Therapy

Some of the most interesting data that we collect as couple and family researchers is linguistic data, whether spoken language (e.g., psychotherapy sessions and clinical interviews) or open-ended responses on questionnaires. By its very nature, linguistic data is unstructured, which is both its strength and weakness. For example, the Oral History Interview (Gottman & Krokoff, 1989) provides a rich description of a couple's current and past functioning, which is highly specific to the couple. The semi-structured format lends broad parameters to the interview's content, but the majority of the data (i.e., the narrative) is generated by the couple. This type of data stands in sharp contrast to forced-choice, questionnaire data, in which item content and response categories are defined by the researcher in advance. Because of this self-report questionnaires limit responses a priori, though analyses of such data are much more straightforward. Conversely, the highly unstructured nature of text data allows a richer source of information – in that participants have more opportunity to reveal information about themselves – but is challenging to analyze. Typically, the analysis of linguistic data is done by applying a behavioral coding system, in either an exploratory, inductive fashion (e.g., qualitative content analysis coding) or via a pre-defined, deductive coding system (Kerig & Baucom, 2004).

While behavioral coding data has yielded many important findings within couple and family research, it also brings its own challenges: a) training raters to use coding systems is time consuming; b) inter-rater reliability can be challenging; c) the actual coding process typically takes a large amount of time; d) for deductive coding systems, one only codes what is specified in their manuals; and e) the portability of a given coding system across research labs can be challenging due to interpretations of coding manuals. In addition to these methodological issues, observational coding methods have also been criticized for unintentionally disguising potentially rich and meaningful distinctions in related interaction behaviors by applying labels that are too broad or sweeping (e.g., Heyman, 2001). Thus, an alternative method for analyzing unstructured text that is highly reliable, portable, and flexible and that quantifies specific aspects of interaction behavior without losing the nuances of families' and couples' private communication systems (i.e., Hooper, Knapp, & Scott, 1981) could be a welcome addition to the family researcher's analytic toolbox.

The current paper presents a novel method (within family psychology) for analyzing text data called topic models (Blei, Ng, & Jordan, 2003; Steyvers & Griffiths, 2006). Topic models provide an alternative, data-driven quantitative method for analyzing text data such as transcripts, journal entries, or other open-ended response questions. Topic models share some similarities with other dimensionality-reduction techniques, such as cluster analysis or principle components analysis (Tabachnick & Fidell, 2007), but are designed specifically for use with text. Similar to these other methods, topic models search for semantic structure without input from the data analyst and hence provide a broad summary of the semantic content as opposed to finding semantic content in pre-specified areas. An advantage of this method is that, in addition to greatly reducing the dimensionality of text, the topics (i.e., word clusters) are typically quite interpretable, which makes them useful for situations in which the goal is not only data reduction but also interpreting and understanding the basic, underlying dimensions of the linguistic data.

In the remainder of the article, we provide an overview of topic models using several illustrative applications to a corpus of linguistic data from a couple therapy trial, comprising transcripts of therapy sessions and semi-structured communication assessments (Christensen et al., 2004). The present article is not intended to be a tutorial in fitting topic models, though supplementary material provides additional technical details and practical information on fitting topic models. Instead, the intent of the present article is to familiarize the reader with this approach and illustrate its potential uses in family psychology. The illustrative applications focus on two broad areas:

1. **Summarization and exploration of transcripts:** By summarizing a therapy session via topics, the model provides a concise overview of the themes in the linguistic data of each therapy session. Topic models assign individual words to specific topics, allowing a researcher to browse through transcripts according to a topic, or locate portions of transcripts that are highly related to specific topics. In addition, a researcher can investigate temporal changes in topics, which can reveal general changes in positive or negative language or sessions with specific content (e.g., particular interventions or important topics such as infidelity or divorce).
2. **Prediction of behavioral codes:** Topic models can also be used in prediction models of behavioral codes or other external outcomes (i.e., external to the text). Topic models summarize semantic information quantitatively and could provide a method for assigning behavioral codes that are semantic in nature. More generally, topic models provide a reduced dimensional description of the semantic content of a corpus that could be useful for predicting outcomes.

To be clear, we do not see topic models as a replacement for existing coding methods but rather as an additional, supplementary tool for analyzing text data – but one that overlaps behavioral coding with a strong semantic basis. Topic models can be used to efficiently and reliably find dimensions of the data that lie beyond the scope of a coding system but also can be trained to locate topics that are associated with existing codes. Thus, they provide an alternative, and in some cases complementary, set of tools for text data relative to behavioral coding.

Overview

Perhaps the easiest way to intuitively understand what a topic model does is to examine the types of output that this model provides.

Figure 1 presents 18 (out of a total of 100) topics that were derived by applying the topic model to the corpus of transcripts from Christensen et al. (2004). (Further detail on the corpus is provided below.) For each topic, the fifteen most probable words are presented in descending order. Reviewing the topics and words in Figure 1, it is clear that the high-probability words within each topic capture semantically related content. Similar to factor analysis, the model provides estimates of how closely each word is associated with a given topic, but it is up to the researcher to discern what this content might be (i.e., how to name the factor or topic). Thus, the first group of topics in the top row of Figure 1 was labeled as “content-based topics” as they tap language that might be discussed during couple therapy, such as family, work, money, and sex.

The second group of topics in Figure 1, labeled “emotion-based topics”, capture features related to specific emotional content, whereas the third set of topics in Figure 1 (i.e., “therapy-related topics”) capture features of therapist-intervention language specific to the two approaches in the Christensen et al. (2004) study (i.e., traditional behavioral couple therapy [TBCT; Jacobson & Margolin, 1979] and integrative behavioral couple therapy [IBCT; Jacobson & Christensen, 1998]). TBCT is a skills-based therapy that focuses on communication training and problem-solving, which is apparent in the topics presented in the bottom row of Figure 1. In contrast, IBCT focuses on identifying couple's relational patterns and underlying emotions. One of the strengths of topic models is that they derive the linguistic structure inductively, without any specific guidance about what type of linguistic content it should find. Thus, a topic model of the couple therapy corpus has identified a number of semantic clusters that are directly relevant to the content and process of couple therapy.

After fitting the model, researchers can then use these topics to summarize and explore the semantic content of a set of therapy sessions, or to make predictions about other aspects of therapy (e.g., behavioral codes or therapy outcomes). Before presenting potential applications of topic models, we first describe the couple therapy corpus in greater detail, provide details about text processing prior to model fitting, and then describe some of the statistical underpinnings of topic models.

Couple Therapy Corpus

The present text data came from a randomized trial of two couple therapies for chronically and stably distressed couples seeking treatment (Christensen et al., 2004). Couples ($N=134$) were randomly assigned to one of two treatments: TBCT or IBCT. TBCT is a skills-based couple therapy that has a didactic style, where the primary focus of the therapist is helping the couple to learn new skills to ameliorate their relationship problems. IBCT is strongly rooted in a behavioral case conceptualization approach, in which the therapist helps the couple to identify core patterns or themes in their relationship and also helps spouses express and respond to vulnerable emotions that are often underlying intense conflict. Couples received up to 26 sessions of therapy and completed self-report assessment batteries prior to therapy, at 13 and 26 weeks after the start of therapy, at the final session of therapy, and several times over five years of post-therapy follow-up. In addition, couples completed structured communication assessments prior to therapy, at the 26 week assessment, and at two years post-therapy. Finally, there was also a small, non-distressed couple comparison group ($n=48$). These couples were selected for having satisfying and functional relationships and completed a single assessment battery, mirroring the pre-therapy assessment of the treatment-seeking couples. Primary treatment outcomes at post-therapy, two years post-therapy, and five years post-therapy are reported elsewhere (Christensen et al., 2004; 2006; 2010) as well as outcomes from observational coding of the communication assessments (Sevier, Eldridge, Jones, Doss, & Christensen, 2008) and therapy sessions (Sevier, 2005).

As described above topic models work with text data, and a supplemental grant from NIMH provided for significant transcription of therapy sessions and communication assessments, which comprise the couple therapy corpus of text used in the present analyses. For therapy sessions, couples were randomly selected for transcription stratified by treatment condition and initial distress. For the 91 couples selected for therapy session transcription, four sessions were transcribed in their entirety: the first session, the next to last session, and then two sessions closest to standardized assessments at 13 and 26 weeks after the start of therapy. For remaining sessions, a randomly selected quarter of the session was transcribed (approximately 13 minutes each). This design allowed a greater number of sessions to have some transcription from each couple, albeit incomplete transcription. Finally, communication assessments at pre-therapy, post-therapy, and two years following therapy were transcribed for all 134 treatment-seeking couples and 48 non-distressed comparison couples. Communication assessments included two, 10-minute problem-solving interactions, in which each partner selected a topic for one of the interactions. In total, the couple therapy corpus contains approximately 6.5 million words (and bracketed expressions, e.g. '[laugh]'), across 1,486 unique therapy sessions and 765 communication assessments.

Text Processing and Topic Models / Latent Dirichlet Allocation

As noted earlier, relative to most other data that family and couple researchers work with, linguistic data is highly unstructured and has high dimensionality.¹ To better understand the dimensionality problem of raw text, a comparison with behavioral coding data may be useful. The communication assessments in the present data were coded using the Couple

Interaction Rating System (CIRS; Heavey, Gill, & Christensen, 2002). The CIRS has 13 codes that were rated for each partner during a communication task. Thus, a 10 minute conversation between spouses is reduced to 26 numeric values through coding with the CIRS, and the overall behavioral coding data could be summarized by a matrix of 26 (items) by 134 (couples) for each communication assessment. On the other hand, the linguistic content of an equivalent corpus of communication assessments will be vastly more complex. Considering each of the individual words spoken as single items, the problem here becomes obvious: The dimensionality of such a model would be enormous as the representation of a corpus would be a matrix, where there could be hundreds up to thousands of sessions (e.g., the couple therapy corpus has 2,251 unique sessions or assessments). Unique word tokens (i.e., a specific instance of a specific word) can often be in the millions. Let's examine how text processing and topic models make this seemingly intractable problem manageable.

Depending on the size of the corpus, the total vocabulary could be quite large, and it is common to take one or two steps to reduce the overall size of the vocabulary. One method of reducing the size of the vocabulary is to stem all of the words. Many different words share the same root. For example, walk, walks, walking, and walked all share the common root "walk". Stemming is an automated method for reducing similar words to their common stem (Porter, 1980). The unstemmed (or raw) couple therapy corpus contained 33,216 unique words, whereas after stemming, the corpus had 22,409 unique words. (Software options for text pre-processing and topic models are considered later.) A second common strategy in pre-processing text data is to remove highly frequent functional words—often referred to as stop-words. For example, the words "a", "of", and "the" are extremely common, but convey relatively little meaning. In our corpus, removal of approximately 600 stop-words² reduced the total number of word tokens from approximately 6.5 to 1.1 million. It is beyond the scope of the present article to give a complete overview of pre-processing but a book-length introduction can be found in Manning and Schütze (1999), or a briefer applied overview using the R statistical software can be found in Feinerer, Hornik, and Meyer (2008). Stemming and stop-word removal serve two key functions: a) they help to reduce the computational burden of fitting topic models (or other text-mining procedures), and b) these procedures enhance the interpretability of topics by removing redundancy (stemming) and words with little content (stop-words).

A final step before topic modeling is to define a "document" in the corpus. In original applications of topic models, documents were quite clear in that the model was often applied to articles or scientific abstracts. Each article or abstract defined a unique document, which is simply a unit of text with some semantic similarity. With spoken language there are a couple possibilities for defining a document. For example, a document could be defined as the full set of words spoken during a therapy session, or could be defined as the set of words spoken by a specific person (e.g., the husband) within a therapy session. In the present analyses, documents were defined as all words spoken by a single individual (i.e., husband, wife, therapist) in a single therapy session or communication assessment; further discussion of defining documents is found below. Once we have defined our set of documents, we can apply a topic model to the corpus. The basic steps are outlined in Figure 2.

A standard method for representing a corpus of documents is using a word-document matrix (WDM). A WDM is simply a frequency (or crosstabs) table in which each of the W unique words in the corpus corresponds to a row, and each of the D unique documents corresponds

¹In the present article we use linguistic data synonymously with text data. More generally, spoken language encompasses both what is said (i.e., semantic content in text) and how it is said (i.e., prosody and tone of spoken language). Although outside the scope of the present article, speech signal processing analyses have examined pitch and other acoustic features of the spoken language in this couple therapy corpus (Black et al., in press).

²The list of stop-words can be obtained from the authors.

to a column. The actual values of each cell in this table are simply counts of the number of times each word occurs in each document. So, if word w shows up five times in document d , the value of the cell $M_{w,d}$ would be equal to five.³ Note that the WDM representation does not capture word-order information, and thus models which use only the WDM as input make what is known as a “bag-of-words” assumption (where the ordering of the words is assumed to be unimportant). Despite the fact that we know that word-order is very important for the particulars of human communication, word order is less critical for deriving the basic semantic dimensions of text. This assumption is extremely common, since it greatly simplifies model complexity (Harris, 1954). Nonetheless, it is important to realize that this is an assumption of the basic topic model, which could affect certain applications. Extensions to the basic topic model presented in the current paper have examined incorporating word order though model and computational complexity increase enormously (see, e.g., Griffiths, Steyvers, Blei, & Tennenbaum, 2005).

Once we have constructed our word-document matrix from the set of documents in the corpus (Figure 2, step 2), we can then apply the topic model to this dataset (Figure 2, step 3). As previously indicated, topic modeling is an unsupervised⁴ machine learning method for finding a set of topics that can be used to summarize a collection of documents. In the technical literature, topic models are often referred to as latent Dirichlet allocation (LDA; Blei et al., 2003), which characterizes the statistical model that is the basis for topic modeling. Given the input of the WDM, the model derives: 1) a set of topics that captures underlying semantic themes in the corpus, 2) a representation of each document in terms of the set of topics, and 3) the assignment of each individual word within the corpus to a specific topic (Figure 2, step 4). Specifically, each topic is modeled as a probability distribution (i.e., a mixture) over words, and each document is modeled as a probability distribution of these topics, where the topics with high probability for a given document capture the semantic themes that are most prevalent within the document. Here in the main text, we avoid technical details underlying the model; additional details about the model, its estimation, and model settings can be found in the supplementary technical appendix.

The matrix of topic-word probabilities can be used to visualize the semantic themes in the corpus (as illustrated in Figure 1). The document-topic matrix can be used to summarize the semantic themes in a specific document (e.g., a transcript from a single therapy session or assessment); for example, one can quickly get a sense of what themes are in a document by looking at just the top few topics that occur in a document. Additionally, the document-topic matrix can be used to create topic model-based regression covariates for additional modeling of outcomes or codes (i.e., just as principal component analysis can be used as a pre-processing step for extracting a lower-dimensional representation to be used in regression modeling, the output of LDA similarly provides a lower-dimensional representation, where the number of features [e.g., covariates in a regression] is equal to the number of topics). The remainder of this paper will focus on the ideas described in this paragraph, using the couple therapy corpus.

We first provide several illustrations of how LDA can be used for data exploration and summarization and then discuss using the output of topic models for discovering associations of topics with behavioral codes, including details on appropriate methods for regression models with topic model data as covariates. Finally, we discuss some simple extensions of the standard LDA procedure, which can be useful for between group

³We note in passing that WDM are typically very sparse matrices in that many of entries are zero. Because of this WDM are stored using a sparse matrix representation in which only non-zero entries are recorded in triplet format (i.e., a three number encoding of word, document, and frequency). This vastly reduces the overall size of the matrix and greatly reduces computing time.

⁴In machine learning parlance, unsupervised methods are those that find structure without additional input, such as cluster analysis, whereas supervised methods are those that are learning associations to a particular outcome, such as regression models

comparisons and ways that behavioral codes can be used directly in the topic model (i.e., the topic generation process is informed by behavioral coding data).

Using Topic Models to Summarize and Explore Text

Following the pre-processing of the text data described earlier (i.e., stemming and stop-word removal), a series of topic models were fit to the couple therapy corpus, varying the number of topics (25, 50, 100, and 200). The number of topics is set by the data analyst when estimating the model, and different numbers of topics may be useful for different tasks (i.e., description and summary vs. predicting outcomes). The supplementary material provides additional details on determining the number of topics. In the current section, we use the 100 topic model to explore the text and examine the semantic themes found by the model, and models with additional topic numbers are used later in the prediction of behavioral codes. Figures 3 and 4 show brief sections of transcript in which the topic model assignment of given words to specific topics is indicated by color-coding and numeric superscript. Note that words that are not assigned to a topic either were stop-words that were removed, or did not cross a threshold for assignment by the model (i.e., each word has a posterior probability from the model of being assigned to a specific topic and in some cases the model cannot make a clear decision). In Figure 3 a therapist is describing some of the basic elements of problem-solving.

This brief section shows that the topic model has located several semantic clusters related to “problem-solving language”, specifically topics 49, 60, and 77. Moreover, each of these three topics is describing specific elements of problem-solving, including defining the problem (topic 60), generating problem solutions (topic 49), and weighing the options (topic 77). At the same time, this section also shows the challenges of similar (or exact) words being used to convey slightly different meanings. The final statement from the therapist is asking whether the wife is having “problems” with the problem-solving model, which also gets assigned to topic 60. Although topic models were developed in part to deal with polysemy (i.e., words with different meanings such as play or bank; see, e.g., Griffiths, Steyvers, & Tenenbaum, 2007), nuanced word usage such as the just noted example can be challenging for the model to discriminate.

Figure 4 shows a section of transcript from a different couple with a notably different focus. This portion of transcript from a couple that is not doing well in therapy shows how the model is locating negative affect words and topics.

The topic model is clearly identifying this portion of transcript as quite negative, with topics on hatred (topic 36), anger / disgust (topic 3), arguments / screaming (topic 16), and divorce (topic 44). Readers will also notice that some word assignments to topics are not readily obvious. For example, truly and guess are assigned to topic 3 as are anger and disgusted. Topic models fit a distribution of topics over documents, meaning that some topics are prevalent in a given document, whereas others are not. Because of this feature of topic models the base-rate of a topic within a document influences the assignment of more ambiguous words in that document. With the present example, the session had strongly negative content, which raises the general probability that words get assigned to negative topics, assuming that the model does not have clear indication to assign it to an alternative topic. However, by looking at the most prevalent words in topic 3, which include anger, angry, and hurt, it is clear that the topic is largely reflecting conflict and negative affect.

The previous two illustrations focused on specific portions of transcripts, which is helpful for understanding how the model handles text at the word-level. It is also possible to use topic models to explore broader semantic themes by summarizing topic model output over entire therapy sessions. For example, the relative prevalence of topics for a single therapy

session portrays which semantic themes were prevalent for that particular session. Moreover, we can use topic model output to examine which topics are prevalent for a particular couple across all sessions. To illustrate this, Figure 5 presents an image plot that shows specific topic distributions over time (i.e., therapy sessions) for a single couple. Each row in this figure corresponds to a single topic, and each column corresponds to a single therapy session. The relative proportion of words assigned to each topic within a given session is represented by the brightness of the corresponding cell; brighter cells indicate that a higher proportion of words were assigned to the topic for that session. The topics have been divided into three groups, relating to: a) intervention language (the top block of topics), b) therapy content / problems (middle block), and c) affective topics (bottom block).⁵

Figure 5 is a representation of how the semantic content in therapy evolved over time for this couple. In examining the top blocks of topics, topic 83 relates to communication training language, and it is most prevalent in sessions 4–8 at which point there is a shift toward problem-solving (topics 6, 46, and 60). The middle block of topics illustrates the particular content that were most central for this couple. As we might predict given that this couple received TBCT, these content topics were discussed in the first few, assessment-oriented sessions and then appear again around the time that the problem-solving intervention is introduced (sessions 9–12). For this particular couple, the wife was a flight attendant and traveled frequently with erratic schedules. Work and finances were notable concerns for them (topics 23, 91, 92). In addition, her schedule plus their young children also meant the spouses were frequently exhausted with limited energy for each other (e.g., topic 100). After effective use of problem-solving on these topics, the final portion of therapy shifted to helping the couple find time for doing enjoyable activities, including planning for a vacation and exercising more (topics 4 and 81). The final set of topics show how the affective content of sessions shifted over time for this couple. Initially, there was notable “frustration” language as they described their relationship (e.g., topic 85). Over time, this topic largely drops out of the couple's transcripts, whereas positive affect as seen in topic 69 becomes relatively more prevalent.

This example shows how the output from topic models can provide an overview of significant themes in a couple's course of therapy. Moreover, these examples focused on specific couples, whereas a similar process could be used with the entire sample, which might highlight particular problems that were prevalent in couples overall and general patterns of intervention language over time for the sample as a whole.

Using Topic Model Output to Predict Behavioral Codes

As noted earlier the most common method for analyzing linguistic data is behavioral coding, in which coding systems define a priori linguistic content that is thought to be important. As an example of how topic models can encode semantic information to be used in predictive models (e.g., regression models), we examine how topic models from the couple therapy corpus can predict behavioral codes. At the outset, we note that this is a specific application of a much more general process of using topic models to quantify semantic information, which is then used to predict external outcomes (i.e., external to the text).

Due to the high-dimensional nature of linguistic data, dimensionality-reduction and feature-selection are common practices in the area of text-based regression modeling. The reason for this is directly tied to our earlier discussion of dimensionality: Suppose that we wanted to

⁵As noted earlier, the model was fit using 100 topics, and thus those presented in Figure 5 are a small subset. Not surprisingly, out of 100 different semantic themes, many are not relevant for any single couple. In generating Figure 5 we began by displaying only those topics that were prevalent in this couple's language above a certain threshold, which then led to an interpretable subset.

use our corpus of therapy transcripts to predict whether a couple will get divorced or not. Even after stemming and stop-word removal, the corpus contains over 20,000 unique words. Using logistic regression or a variant thereof, this would require estimating coefficients for each of the over 20,000 unique terms in our corpus. Given that we have data for only 91 couples (i.e., couples with transcripts and outcomes), this model would be greatly over-parameterized, which presents a problem for developing a reasonable regression model.⁶ A common solution to this issue of over-parameterization is to use dimensionality-reduction, such as principal components analysis, to reduce the number of coefficients in the model. In a similar vein, topic models provide a lower-dimensional representation of the data; each document (i.e., session or assessment) can be summarized by the T topics, rather than by counts for each unique term in the corpus, which both simplifies the regression modeling but also increases interpretability by using topics as predictors.

A subset of therapy sessions in the Christensen et al. (2004) study was coded using the Couple Therapy In-Session Behavior Rating System (Sevier & Christensen, 2002). This coding system rates each partner separately on 17 behavioral items, using a 1–9 ordinal scale. The codes comprise four subscales: constructive change behaviors, acceptance promotion behaviors, positive behaviors, and negative behaviors. The models described below were fit to each of the 17 items, though for parsimony, we focus on four items, one each from the four subscales: constructive problem solving skills (CPS), descriptive / non-blaming discussion (DIS), positive emotion (PEM), and blame (BLA). There were a total of 1,470 sets of ratings for individuals with matching transcription data available for analysis. The original, ordinal scaling was converted to a binary outcome, taking the top and bottom 20% of the ratings associated with each behavioral code. The rationale for this procedure is that the ratings in the tails of the distributions are most informative for distinguishing between the high and low end of the rating scale, and allows for a balanced binary outcome (i.e., 50% accuracy represents chance). Furthermore, as noted by Georgiou et al. (2011), the distributions of most of the behavioral codes in this dataset tend to be unimodal, with the majority of coded sessions taking on intermediate values for most codes. The distributions of these code values—as well as the variance across individual coders—indicates that for the majority of codes, only sessions for which codes take on extreme scores (e.g., the top and bottom 20% of sessions), are reliably discriminated by human coders. Therefore, our focus is on an approach of utilizing only the extreme scores in our classification task, which has been employed in previous studies involving modeling behavioral codes (Georgiou et al., 2011; Black et al., in press). Based on this procedure, 588 therapy sessions (40% of the 1,470 therapy sessions total) were used for model estimation and testing. For comparison purposes, we also investigated the approach where no data was removed, i.e., in which the ratings were converted to a binary outcome by taking the top and bottom 50% of ratings.

Coding outcomes were predicted from topic model output using topic models with 25, 50, 100, or 200 topics. For each of these topic models, sessions were summarized by the proportion of words assigned to each of the topics. For example, each session from the 25 topic model will have a vector of 25 numbers describing how prevalent each of the 25 topics was in that particular session. Thus, the number of topics defines the number of covariates in each of the predictive models. Regression models with 100 or more covariates are not that common in couple and family research, and blindly fitting a standard logistic regression could very well lead to poor-fitting models. Thus, we provide details on how the current

⁶Some readers may wonder whether such a model would even be possible to estimate. Although not currently common in family science, machine learning methods have been developed to deal with situations in which the number of potential predictors is larger (and sometimes far larger) than the number of individuals or samples, which is common in some genetic studies, for example. See Hastie, Tibshirani, & Friedman (2009) for an overview.

models guarded against over-fitting, an issue that is standard to address in the machine learning literature.

For each behavioral coding item, a sparse logistic regression model was fit (Krishnapuram, Figueiredo, Carin, & Hartemink, 2005). The sparse logistic regression model is based on logistic regression but adds a penalty term for the regression coefficients such that the regression ends up with only a subset of the covariates associated with non-zero coefficients. This procedure is useful in situations where a large number of covariates are used and improves generalization performance (Hastie et al., 2009). The performance of the sparse logistic regression model was assessed by a simple accuracy measure of percentage of codes correctly classified. Because the base rates of the two binary outcomes were equal, chance performance for this model is 50%. Cross-validation was used to assess performance, a standard approach to evaluate the predictive ability of regression models and other supervised learning procedures (Hastie et al., 2009). With the couple therapy corpus a “leave one couple out” cross-validation was used. All of the data from a single couple was removed (or withheld), and the sparse logistic regression was fit to the remaining couples (i.e., all but one couple); this is often called the training set. The fitted regression model was then used to predict the behavioral code for the one couple that was withheld. This procedure was then iteratively applied, leaving out each couple in the sample, and the overall prediction accuracy was averaged across all the individual predictions. Using this type of cross-validation provides strong evidence for generalizability of the model to predict new couples. 7 The difference in prediction between cross-validated and non cross-validated results is an estimate of the degree of over-fitting, somewhat similar to the difference between R^2 and adjusted R^2 . For example, a model with a large number of regressors is expected to perform quite well on the sample data but can potentially show poor generalization performance as assessed through cross-validation.

Figure 6 shows the prediction results from using sparse logistic regression to predict behavioral codes based on topic modeling covariates. The results are shown for the four selected behavioral codes as well as aggregated across the 4 codes (“ALL”). Different lines show the results for leave one couple out cross-validation (red line) and results with no cross validation (black line). Within an individual code, each line shows from left to right, the prediction accuracy based on regression covariates from topic models with 25, 50, 100, and 200 topics. Overall, results show that the topic models predict the binarized codes with between 65% and 70% accuracy and that results vary by the number of topics extracted. There is some consistency across codes that covariates based on 200 topics yield inferior predictions to those with 25, 50, or 100 topics. This makes some substantive sense in that the codes are characterizing broad content that would be applicable to many couples (e.g., positive emotion). A topic model with 200 topics would have many idiosyncratic topics that would be highly specific to individual couples or small subsets of couples. As discussed in the technical appendix, this general procedure of iteratively fitting cross-validated regression models to an external criterion while varying the number of topics is one route to determining an appropriate number of topics to extract. At least with respect to predicting the current behavioral codes, these results would suggest that 25–50 underlying semantic dimensions are optimal. In addition, the results show that models without cross-validation are too optimistic and such prediction models would likely not generalize.

⁷We also examined an alternative cross-validation approach. In this second approach 10-fold cross-validation was applied at the session level by leaving out a random 10% of sessions in the model estimation procedure and measuring prediction accuracy on the held out sessions (this process of leaving out 10% of sessions was repeated 10 times on non-overlapping sets of held-out sessions). In this approach, we test the model’s ability to predict codes in novel sessions but from the same couples. As might be expected, prediction accuracy was somewhat better using this approach relative to the leave one couple out approach, as the model has information about the couples it is trying to predict.

We also investigated whether it makes any difference to use only the top and bottom 20% of ratings when constructing binary outcomes, or to use the full dataset. In the latter case, we created binary outcomes based on the top and bottom 50% of ratings. Performance remained fairly consistent, independent of how much data was included: in the 20%-tails procedure, average prediction accuracy across all 17 codes was 58.9%, with performance significantly better than chance on 13 out of 17 of the total codes as measured using a binomial exact test. In the full-data condition (50%-tails), average prediction accuracy across all codes was 56.3%, with performance significantly better than chance on 15 out of 17 codes. Therefore, our prediction results did not change much by focusing on the extremes of the ratings.

Finally, we investigated whether the resulting coefficients from the sparse logistic regression make substantive sense.

Figure 7 presents a summary of the results of these models. For each code, the six topics with the largest regression coefficients are shown in rank order, and the top 10 words for each topic are provided. The bar plot for each code shows the relative strength of the regression coefficient. Across the four outcomes, the most predictive topics provide an interpretable set of features that make intuitive sense in terms of the codes with which they are associated. For example, the top three topics that are most predictive of high scores on CPS all clearly relate directly to the issue of problem-solving. The first three topics have certain high-probability words such as pros, cons, and brainstorming that capture specific language associated with problem-solving interventions, whereas the remaining three topics appear to capture linguistic content that may be a focus on problem-solving interventions (or communication strategies to use in the course of problem-solving). These final three topics also demonstrate the utility of topic models for preserving the nuances of couples' language within a quantitative framework. From these results, it is clear that sex and communication were two of the topics most frequently addressed within a problem-solving framework. Though it may not be surprising to anyone who has worked with couples clinically that sex and communication were common topics of discussion, topic models identify the presence of these themes in a manner that allows themes to be linked with other behaviors of interest. Because topic models are not confined to predetermined codes, we could use the present finding to ask more inductive research questions such as, "What are the most common themes addressed using problem-solving training in successful and unsuccessful couple therapy cases?" and do so in a quantitatively rigorous manner.

For the two codes relating to emotional content, PEM and BLA, the regression coefficients are clearly picking up on topics that specifically capture affective content in the language. For PEM, there are high-probability words with positive emotional content (e.g., good, nice), but also bracketed-terms that capture positive emotional expressions ([laugh], [laughter]). For the BLA code, the topics capture words at the other end of the emotional spectrum. The most highly weighted topic seems to capture language that would arise during heated conflicts, and other topics include swearing, trust, and also children and money, presumably therapy topics around which blame can arise. As with the associations between sex, communication and CPS, these results help to contextualize blame language in a manner that is very difficult to achieve with observational coding. For example, many of the words in the swearing topic are dismissive statements that are not overtly critical but that function to create emotional distance from a spouse in much the same way as the overtly angry and critical words in the first topic. Note that, although BLA contrasts strongly with PEM, these two codes are not exact opposites; although we would expect that blame would tend to co-occur with negative emotions, the notion of blame is more subtle than simply "negative emotions". Although it is difficult to be certain whether the topic model-based regression is picking up on this distinction, there is some evidence that it is. Specifically, whereas the topics with large coefficients for PEM all seem to capture topics with positive

emotional content, the most predictive topics for BLA are not all directly related to negative emotional content but also include substantive topics such as children and money.

This example was intended to illustrate how topic models might be used in predictive models of external outcomes. In the present case, the external outcomes were behavioral codes, and the results show that topic models have strong concordance with some behavioral codes. The ability of topic models to predict behavioral codes will be directly related to how strongly semantic a given code is. For example, another code in the current coding data is “withdraw.” This code captures a range of behaviors that are primarily non-verbal (i.e., not engaging nor responding to partner), and the prediction accuracy of the regression model reflects this as it is no better than chance. Thus, for behavioral codes with strong semantic content, topic models and regression procedures just described could be one route to developing a model-based coding procedure. That is, given the availability of some coding data to train the model, procedures such as those described above could be used to estimate codes for uncoded transcripts. In addition, the general process described above could be used with other criterion variables, such as therapy outcomes. These types of models could be used to characterize linguistic content prior to or during therapy that is associated with good (or poor) outcomes.

Limitations, Additional Resources, and Future Directions

Topic models have a notable, practical limitation: They require transcripts. There was significant cost and effort required to generate the couple therapy corpus. However, it seems likely that this limitation will lessen over time. More and more data are being collected electronically, and the days of “paper and pencil” measures are quickly fading. Thus, open-ended response data is often collected electronically, without requiring additional transcription. For spoken language, recording quality is consistently improving, as is speech recognition software. Thus, we are guardedly optimistic that the transcription barrier to using topic models is lowering. A different limitation is that topic models are fundamentally lexical and semantic models, and they do not capture linguistic tone for transcripts of spoken language nor nonverbal behavior that might be evident in a video recording (see, e.g., Black et al., in press; Ramseyer & Tschacher, 2011).

Our goal in this article was to provide an applied introduction to topic models, which provide a quantitative method for analyzing text that is new to the field of couple and family research. As with any introduction paper, it is impossible to go into detail on all of the relevant issues relating to the topic models. In the supplementary appendix of this paper, we provide further technical details and references for those interested in understanding the underlying mathematical details of the topic model. Here, we provide some additional comments relating to variants of the topic model that may be useful for future directions of research, as well as resources for those interested in learning more about the model or in using the model themselves.

The model employed in this paper is the standard LDA or topic model (Blei et al., 2003). However, the flexible nature of this model has led to the development of numerous variants and extensions of the model (e.g., see: Blei, 2011; Griffiths et al., 2007) and topic modeling now comprises a large and diverse field of research. Many of the more recent extensions of topic models could be extremely useful in the context of family psychology. For example, several topic models have incorporated both text and additional data such as labels or annotations applied to documents (e.g., Blei & McAuliffe, 2007; Mimno & McCallum, 2008); similar approaches could be beneficial in the context of clinical psychology, and could be used for exploring joint models of both textual data and the wide array of additional data collected in clinical work (e.g., behavioral codes, questionnaire data, etc.).

Furthermore, whereas the results presented in this paper used topics derived without any knowledge of the behavioral codes, variants of topic models could incorporate this supplementary clinical data directly into the generative process for topics. Not only is this useful in that it can be used to associate linguistic content with specific codes (Ramage et al., 2009), but this class of topic models is capable of achieving highly-competitive prediction performance (Rubin et al., in press). This specific variant of topic models may serve as excellent models for exploring areas such as automated behavioral coding.

Resources on working with text and pre-processing text prior to running topic models can be found in Manning and Schutze (1999) and Feinerer et al. (2008). At the present time, topic models are easily accessible in MATLAB via the topic modeling toolbox (Steyvers & Griffiths, 2011) in R via the tm (for text-mining; Feinerer et al., 2008) and topicmodels packages (Gruen & Hornik, 2011), and in a variety of other languages.⁸

Concluding Remarks

Albert Einstein famously said: “If atoms could talk, I would surely listen.” Couple and family researchers have tried to listen to the talk that constitutes so much of close relationships. However, they have not always known how to make sense of that talk and have often resorted to the constraints of forced choice questionnaires or pre-determined coding categories as a method of studying relationships. Topic models represent a way to avoid these constraints and a method for couple and family researchers to use open-ended text data in its full richness. They are one way in which qualitative data can be used in quantitative modeling and offer added flexibility over coding systems. Moreover, the initial applications described here show that they can detect important linguistic processes in interaction data both at the level of what is discussed (i.e., content) as well as how it is discussed (i.e., emotion). We hope that these methods might enable couple and family researchers to more directly model linguistic hypotheses and therapy mechanisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Zac Imel, Scott Baldwin, and three anonymous reviewers for helpful feedback that improved the manuscript in numerous ways. The couple therapy data used throughout the manuscript were collected via an NIMH grant awarded to Andrew Christensen at the University of California, Los Angeles (MH56223) and Neil S. Jacobson at the University of Washington (MH56165). David Atkins, Tim Rubin, and Mark Steyver's time was supported in part by NIAAA grant AA018673 (Atkins / Steyvers, co-PIs).

References

- Black M, Katsamanis A, Baucom BR, Lee CC, Lammert AC, Christensen A, Georgiou PG, Narayanan SS. Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech Communication*. (in press).
- Blei D, Ng AY, Jordan MI. Latent Dirichlet allocation. *Journal of Machine Learning Research*. 2003; 3:993–1022.
- Blei, D.; McAuliffe, J. *Neural Information Processing Systems (NIPS)*. 2007. Supervised topic models. 2007

⁸David Blei's topic-modeling webpage (<http://www.cs.princeton.edu/~blei/topicmodeling.html>) is also an excellent resource, and contains a number of introductory/review papers on topic models, links to web-based topic-modeling applications and browsers, as well as a number of implementations of various topic models in R, C and Python.

- Blei D. Introduction to probabilistic topic models. Communications of the ACM. 2011 Available at: <http://www.cs.princeton.edu/~blei/papers/Blei2011.pdf>.
- Christensen A, Atkins DC, Berns S, Wheeler J, Baucom DH, Simpson LE. Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology*. 2004; 72:176–191. [PubMed: 15065953]
- Christensen A, Atkins DC, Yi J, Baucom DH, George WH. Couple and individual adjustment for two years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy. *Journal of Consulting and Clinical Psychology*. 2006; 74:1180–1191. [PubMed: 17154747]
- Christensen A, Atkins DC, Baucom B, Yi J. Couple and individual adjustment for 5 years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy. *Journal of Consulting and Clinical Psychology*. 2010; 78:225–235. [PubMed: 20350033]
- Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. *Journal of Statistical Software*. 2008; 25:1–54.
- Georgiou, PG.; Black, MP.; Lammert, A.; Baucom, B.; Narayanan, SS. Proceedings of Affective Computing and Intelligent Interaction (ACII). Memphis, TN: Lecture Notes in Computer Science; 2011. "That's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?. 2011
- Gottman JM, Krokoff LJ. The relationship between marital interaction and marital satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology*. 1989; 57:47–52. [PubMed: 2487031]
- Griffiths, TL.; Steyvers, M.; Blei, DM.; Tenenbaum, JB. Integrating Topics and Syntax. In: Saul, LK., et al., editors. *Advances in Neural Information Processing Systems*. Vol. 17. MIT Press; 2005. p. 537–544.
- Griffiths TL, Steyvers M, Tenenbaum JBT. Topics in semantic representation. *Psychological Review*. 2007; 114:211–244. [PubMed: 17500626]
- Gruen B, Hornik K. topicmodels: An R package for fitting topic models. *Journal of Statistical Software*. 2011; 40:1–30.
- Harris Z. Distributional structure. *Word*. 1954; 10:146–162.
- Hastie, T.; Tibshirani, R.; Friedman, J. *Elements of statistical learning: Data mining, inference, and prediction*. 2nd Ed. New York: Springer; 2009.
- Heavey, C.; Gill, D.; Christensen, A. *Couples Interaction Rating System – 2nd Edition (CIRS-2)*. Los Angeles: Unpublished manuscript, University of California; 2002.
- Heyman RE. Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment*. 2001; 13:5–35. [PubMed: 11281039]
- Hooper R, Knapp ML, Scott L. Couples' personal idioms: Exploring intimate talk. *Journal of Communication*. 1981; 31:23–33. [PubMed: 7204626]
- Jacobson, NS.; Christensen, A. *Acceptance and change in couple therapy: A therapist's guide to transforming relationships*. New York: Norton; 1998.
- Jacobson, NS.; Margolin, G. *Marital therapy: Strategies based on social learning and behavior exchange principles*. New York: Brunner/Mazel; 1979.
- Kerig, PK.; Baucom, DH., editors. *Couple observational coding systems*. Mahwah, NJ: Erlbaum; 2004.
- Krishnapuram B, Figueiredo M, Carin L, Hartemink A. Sparse Multinomial Logistic Regression: Fast Algorithms and Generalization Bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. 2005; 27:957–968.
- Mimno, D.; McCallum, A. Topic models conditioned on arbitrary features with Dirichlet-multinomial regression; Conference on Uncertainty in Artificial Intelligence; 2008.
- Manning, C.; Schütze, H. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press; 1999.
- Porter MF. An algorithm for suffix stripping. *Program*. 1980; 14:130–137.

- Ramage, D.; Hall, D.; Nallapati, R.; Manning, CD. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-Labeled Corpora; Conference on Empirical Methods in Natural Language Processing (EMNLP); 2009.
- Ramseyer F, Tschacher W. Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*. 2011; 79:284–295. [PubMed: 21639608]
- Rubin TN, Chambers A, Smyth P, Steyvers M. Statistical topic models for multi-label document classification. *Machine Learning*. (in press).
- Sevier M, Eldridge K, Jones J, Doss B, Christensen A. Observed changes in communication during Traditional and Integrative Behavioral Couple Therapy. *Behavior Therapy*. 2008; 39:137–150. [PubMed: 18502247]
- Sevier, M. Unpublished dissertation. 2005. Client change processes in traditional behavioral couple therapy and integrative behavioral couple therapy: An observational study of in-session spousal behavior.
- Sevier, M.; Christensen, A. Couple Therapy In-Session Behavior Rating System. Los Angeles: University of California; 2002. Unpublished manuscript
- Steyvers, M.; Griffiths, T. Probabilistic topic models. In: Landauer, T.; McNamara, D.; Dennis, S.; Kintsch, W., editors. *Latent Semantic Analysis: A Road to Meaning*. Mahwah, NJ: Erlbaum; 2006.
- Steyvers, M.; Griffiths, T. Matlab topic modeling toolbox version 1.4. URL: http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm. [Software]. 2011.
- Tabachnick, BG.; Fidell, LS. *Using Multivariate Statistics*. 5th Ed. Boston: Allyn and Bacon; 2007.

CONTENT-BASED TOPICS									
Family		Finances		Relationships		Sex		Transportation	
Topic 64	.011	Topic 23	.011	Topic 50	.014	Topic 63	.010	Topic 1	.008
mom	.096	money	.104	married	.080	sex	.131	car	.167
mother	.060	dollars	.054	together	.053	sexual	.068	drive	.118
dad	.056	buy	.050	relationship	.045	love	.026	park	.026
sister	.040	hundred	.043	date	.044	part	.021	down	.018
brother	.030	card	.031	live	.036	interesting	.018	street	.017
call	.028	thousand	.021	met	.034	initially	.017	turn	.016
day	.027	bought	.018	move	.028	touch	.016	traffic	.013
father	.023	credit	.017	attracted	.024	back	.016	accident	.013
deal	.018	pay	.015	remember	.023	desire	.016	home	.013
live	.014	fifty	.015	marriage	.022	physical	.015	bus	.012
parents	.014	cost	.014	thought	.018	intimacy	.014	drove	.010
down	.014	give	.014	months	.017	life	.013	hours	.010
care	.013	car	.012	two	.016	talk	.013	direction	.010
stuff	.013	expensive	.012	decided	.015	relationship	.011	freeway	.009
house	.012	five	.012	pretty	.015	pleasure	.010	walk	.008
Topic 91	.010	job	.197	work	.139	career	.027	money	.024
day	.022	company	.019	support	.019	people	.018	situation	.014
hours	.013	week	.013	positive	.012	happy	.011	business	.011
interview	.010								
EMOTION-BASED TOPICS									
Negative Emotional Content					Positive Emotional Content				
Topic 3	.009	Topic 46	.008	Topic 79	.010	Topic 14	.010	Topic 95	.007
angry	.170	give	.023	upset	.113	good	.039	[laugh]	.383
anger	.072	shit	.021	back	.030	thought	.036	guess	.031
hurt	.047	pissed	.021	mad	.030	[laughing]	.031	good	.017
frustrated	.037	point	.019	temper	.024	pretty	.029	work	.017
trying	.032	whatever	.019	talk	.021	people	.025	thank	.012
upset	.027	man	.018	crying	.020	talk	.020	give	.012
mad	.020	fuck	.017	angry	.018	part	.019	wow	.012
point	.018	care	.016	sorry	.018	summer	.017	definitely	.012
sad	.016	god	.016	understand	.017	enjoy	.017	[laughing]	.010
emotional	.016	black	.015	fact	.015	remember	.016	back	.010
part	.016	problem	.015	late	.015	fun	.016	obviously	.010
felt	.014	fine	.015	apologize	.013	u	.016	you'll	.010
whatever	.014	walk	.014	fine	.012	nice	.014	hard	.009
express	.014	white	.013	ready	.012	great	.014	relax	.009
respond	.014	cannot	.013	reason	.012	vacation	.014	[all laugh]	.009
good	.101	nice	.078	thought	.070	felt	.048	appreciate	.033
week	.031	remember	.026	day	.026	couple	.020	notice	.019
work	.019	great	.019	pretty	.017	thank	.016	realize	.015
THERAPY-RELATED TOPICS									
TBCT					IBCT				
Topic 60	.016	Topic 49	.013	Topic 83	.014	Topic 90	.010	Topic 53	.012
problem	.262	solution	.088	listen	.072	pattern	.036	felt	.107
solving	.092	down	.067	communicatio	.053	point	.034	love	.056
issue	.049	write	.062	paraphrase	.030	two	.032	hurt	.052
solution	.035	idea	.051	give	.029	down	.027	care	.052
work	.025	pros	.038	person	.027	become	.023	talk	.049
talk	.022	cons	.037	talk	.027	part	.023	bad	.035
two	.022	brainstorming	.034	trying	.026	conflict	.019	connection	.024
define	.020	good	.025	level	.023	response	.019	close	.018
part	.019	two	.024	speaker	.022	process	.016	heart	.017
definitely	.018	problem	.023	practice	.019	recognize	.016	pain	.014
brainstorming	.015	possible	.021	point	.018	situation	.016	hard	.013
trying	.014	trying	.018	editing	.017	deal	.015	wish	.012
communicatio	.014	discuss	.017	x	.017	withdraw	.015	share	.012
process	.013	talk	.017	floor	.016	style	.014	words	.011
discuss	.012	agree	.016	summarize	.015	change	.013	acknowledge	.010
talk	.047	control	.041	emotional	.040	fear	.029	anxiety	.028
afraid	.026	express	.021	wondering	.021	two	.021	back	.020
anxious	.020	guess	.019	part	.019	hard	.019	scared	.018

Figure 1.

Examples of topics from a topic model applied to couple therapy transcripts. The fifteen most likely words for each topic are shown. Topics have been organized into groups and labeled to aid interpretation. Note that the labels (i.e., in bold typeface) are interpretations of the topics provided by the authors and are not automatically learned by the model.

Step 1

Assemble collection
of documents

Step 2

Create word-
document matrix

Step 3

Run topic model

Step 4

Analyze output from model

	Topic 1	Topic 2
Doc1	70%	30%
Doc2	50%	50%
Doc3	20%	80%

Probability of topics in documents

Topic 1

Probability of words in topics

Topic 2

Doc1

Assignment of word tokens
in a document to topics

Figure 2.

Overview flowchart of how topic models are applied to a corpus of documents (or, in the present example, therapy transcripts).

THERAPIST: ...and the model⁷⁷ is one in which you define⁶⁰ the problem⁶⁰, you brainstorm⁴⁹ ideas⁴⁹, you talk⁷⁷ about the pros⁴⁹ and cons⁴⁹ of the options⁷⁷, and then the last part⁶⁰ of it, or the next to last part⁶⁰ of it is to integrate⁷⁷ the ideas⁴⁹ that you decided⁷⁷ were good⁴⁹ ideas⁴⁹ into a strategy⁷⁷ that you're going to um ... that doesn't mean that we know for sure it is going to work¹⁸. In fact¹⁹ that is the reason⁷⁷ why the last stage⁶⁰ of the model⁷⁷ is to set specified⁴⁹ times that we should reevaluate⁷⁷ and see how your solution⁴⁹ is working¹⁸ or not working¹⁸ and you adjust¹⁸ it accordingly⁴. Right? That was the model⁷⁷, so um, are you um, contemplating⁹⁷ some problems⁶⁰ with the model⁷⁷ or are you um, suggesting⁷⁷ some problems⁶⁰ with the items⁴⁹ that we went through with that, that hasn't⁸⁸ been articulated⁷⁷ yet?

WIFE: Well, I mean I'm, I'm looking at the problem⁶⁰ and the potential⁶⁰ solutions⁶⁰ and, short¹² of me doing something, then nothing will change²².

Figure 3.

Transcript focusing on therapist problem-solving language showing topic assignment for individual words via highlighting and superscripting. Words not assigned to a topic were either stop-words that were removed prior to topic modeling, or did not cross a threshold for posterior probability of assignment (i.e., the model is uncertain about assigning to a specific topic).

WIFE: Well, I don't hate³⁶ him when we get ... not this time, but when we get into those arguments¹⁶ i hate³⁶ him truly³ i mean just before [du]...

THERAPIST: [du] you are really very angry³ with him

WIFE: No not just angry³. Before we started coming it is so much anger³ that i hate³⁶ him. I mean I was ready⁴⁴ for a divorce⁴⁴ right away. I just hate³⁶ him so much we were arguing¹⁶ three times a day⁹⁸. And the days⁹⁸ that he was off work⁴⁴ five⁶⁸ times a day⁹⁸ ...

THERAPIST: Disgusted³ ?

WIFE: Right like being so angry³ you know to each other and scream¹⁶ and ...

THERAPIST: Your face³ looks more like you are disgusted³ I guess³

WIFE: Yeah, yeah, I was just sick²⁷ with that then ... I say I just want a divorce⁴⁴ you know this time was not that, that bad²⁷.

Figure 4.

Transcript focusing on negative affect language showing topic assignment for individual words via highlighting and superscripting. Words not assigned to a topic were either stop-words that were removed prior to topic modeling, or did not cross a threshold for posterior probability of assignment (i.e., the model is uncertain about assigning to a specific topic).

Figure 5.

Image plot of the proportion of words from couple therapy transcripts assigned to 17 topics for a particular couple. Rows of the plot are defined by topics, and columns are defined by session number. Topic proportions are presented in grayscale, and lighter shades indicate higher topic proportions. The color mapping has been scaled for easier interpretation.

Figure 6.

Accuracy in predicting four behavioral codes (CPS = constructive problem-solving; DIS = descriptive non-blaming discussions; PEM = positive emotion; BLA = blame), as well as average prediction of four codes (ALL), based on sparse logistic regression models. Results are presented by the number of topics, type of cross-validation analysis, and behavioral code. For each line, the four points from left to right indicate the results for a topic model with 25, 50, 100, and 200 topics. Accuracy is based on percentage of sessions where the model correctly predicts that the behavioral code is associated with a rating in the bottom or top 20% of ratings.

Figure 7.

The six topics that are most predictive for high scores on four behavioral codes of couple therapy transcripts based on logistic regression models (CPS = constructive problem-solving; DIS = descriptive non-blaming discussions; PEM = positive emotion; BLA = blame). For each topic, the ten most likely words are shown alongside the β weights from the logistic regression.