

# Finding Your Voice: The Linguistic Development of Mental Health Counselors

Justine Zhang

Cornell University

jz727@cornell.edu

Robert Filbin

Crisis Text Line

bob@crisistextline.org

Christine Morrison

Crisis Text Line

christine@crisistextline.org

Jaclyn Weiser

Crisis Text Line

jweiser@crisistextline.org

Cristian Danescu-Niculescu-Mizil \*

Cornell University

cristian@cs.cornell.edu

## Abstract

Mental health counseling is an enterprise with profound societal importance where conversations play a primary role. In order to acquire the conversational skills needed to face a challenging range of situations, mental health counselors must rely on training and on continued experience with actual clients. However, in the absence of large scale longitudinal studies, the nature and significance of this developmental process remain unclear. For example, prior literature suggests that experience might not translate into consequential changes in counselor behavior. This has led some to even argue that counseling is a profession without expertise.

In this work, we develop a computational framework to quantify the extent to which individuals change their linguistic behavior with experience and to study the nature of this evolution. We use our framework to conduct a large longitudinal study of mental health counseling conversations, tracking over 3,400 counselors across their tenure. We reveal that overall, counselors do indeed change their conversational behavior to become more diverse across interactions, developing an individual voice that distinguishes them from other counselors. Furthermore, a finer-grained investigation shows that the rate and nature of this diversification vary across functionally different conversational components.

## 1 Introduction

Conversations are central to many human endeavors and professions, from academic advising, to business negotiations, to customer service, to mental health counseling. The choice of conversational language has been shown to correlate with a broad spectrum of consequential

outcomes such as success in persuasion (Habernal and Gurevych, 2016; Tan et al., 2016; Zhang et al., 2016; Packard et al., 2018), in team performance (Yang et al., 2015; Niculae and Danescu-Niculescu-Mizil, 2016), or in social support (Atkins et al., 2014; Althoff et al., 2016; Choudhury and Kiciman, 2017). However, surprisingly little is known about how adults develop conversational skills. Understanding this challenging process has the potential to help with tracking and fostering behavioral development in conversation-focused endeavors.

In this work, we address this knowledge gap by examining how individuals change their conversational language in a domain with profound societal importance, where conversations play a primary role: mental health counseling. Counselors in this setting face a daunting task: through conversations, they need to empathetically respond to and support clients undergoing psychological distress (Althoff et al., 2016).

Initial training can orient counselors towards the principles of counseling conversations, but cannot cover the broad range of situations they will inevitably encounter. Practical experience, on the other hand, comes with scarce direct feedback: it is hard for counselors to gauge whether a client was positively affected by a conversation. Perhaps as a result, prior studies find no consequential change in counselor behavior or effectiveness beyond their initial training (Dawes, 1996; Hill et al., 2015; Goldberg et al., 2016). This lack of evidence has led some to even argue that psychological counseling is a “profession without any expertise” (Tracey et al., 2014).

However, even in the absence of feedback, multiple forces could lead counselors to change their behavior with experience. First, practice exposes—and perhaps familiarizes—the counselor to a wide range of client situations, far be-

\*Corresponding senior author.

yond what training can encompass. This could enable them to increasingly adapt to the situations they encounter, differentiating their behavior in one conversation from their behavior in another (*within-counselor diversification*).

Second, counselors come into the domain with individual personalities and conversational styles. A strict adherence to training materials can lead to unnatural and robotic conversations, which can be taxing on the counselors (Orlinsky and Rønnestad, 2005) and detrimental to the client (Henry et al., 1993; Castonguay et al., 1996). A counselor might therefore seek to adapt training principles to their own personality, finding a voice that distinguishes them from other counselors (*between-counselor diversification*).

However, change is far from guaranteed, and several forces potentially counteract the two vectors of change outlined above. The challenging nature and high (sometimes literally life-and-death) stakes of the conversations might lead to an overly cautious approach, deterring counselors from departing from language exemplified in the training material and instead leading them to seek “security in a restrictive practice routine” (Skovholt et al., 1997). The lack of feedback might exacerbate this stagnation and inflexibility, since counselors have no direct way to assess the potential effects of behavioral changes (Shanteau, 1992; Kahneman and Klein, 2009).

In this work, we design a computational methodology to quantitatively track systematic changes along the two dimensions of within- and between-counselor linguistic diversification. Our framework additionally exposes the rate at which these changes happen across individuals and highlights particular functional components of their language where this change is especially salient.

We apply this framework to conduct a large-scale longitudinal study of mental health counseling conversations on Crisis Text Line, a text-based crisis counseling platform,<sup>1</sup> tracking the conversational behavior of over 3,400 counselors across their tenure. Our study reveals that overall, counselors do indeed systematically change their conversational behavior, becoming more linguistically diverse across their own conversations as well as among each other. Furthermore, this diversification process advances at different rates

<sup>1</sup>This study was done in collaboration with Crisis Text Line, using anonymized data and following IRB approval. Details of the data access are included in Section 3.

in different functional components of the conversation; in particular, within-counselor diversification is amplified in components concerning the client’s specific problems. This suggests that counselors accumulate domain fluency, perhaps enabling them to better address the particular situations they encounter.

We complement this intuition by examining how counselor vocabulary shifts with experience: which words increase in usage or fall out of favor as counselors gain experience? We find that as they advance in tenure, counselors adopt more specialized as well as more colloquial language, while slowly abandoning language from the training material.

Overall, these results provide the first evidence of a systematic development of mental health counselors with experience, thus contributing to open questions about expertise from the counseling and psychotherapy literature. From a practical standpoint, our framework could aid counseling platforms to automatically identify counselors that stagnate. Additionally, by highlighting conversational practices that require more experience to grasp, this type of understanding can inform sustained training programs across a counselor’s career (Tracey et al., 2014).

More broadly, our work takes initial steps in understanding how individuals develop their conversational behavior. Beyond mental health counseling, this methodology can potentially provide insights into other domains where having conversations is crucial but potentially hard to teach, such as academic advising and customer service.

## 2 Further Related Work

Our work draws on prior literature concerning language use in conversational settings like counseling, as well as on studies of linguistic change.

**Language in the mental health domain.** Prior literature has underlined the importance of linguistic choices and conversational behavior in mental health-related contexts. Qualitative (Labov and Fanshel, 1977; Miller and Mount, 2001; Catley et al., 2006; Gaume et al., 2010) and computational (Atkins et al., 2014; Tanana et al., 2015, 2016; Althoff et al., 2016; Pérez-Rosas et al., 2018) studies of dialogic interactions in counseling and psychotherapy have highlighted the potential benefits of conversational skills—such as a counselor’s ability to reflect on a client’s par-

ticular concerns. Other work has considered a broader range of settings such as online forums, exploring how psychological support is sought and provided (Choudhury and Kiciman, 2017; Ernala et al., 2017; Yates et al., 2017; Yang et al., 2017, 2019; Pruksachatkun et al., 2019). While these studies have focused on linguistic behaviors in the scope of single interactions, we seek to understand how these behaviors develop over time.

**Linguistic behavior in conversations.** Our work relates more broadly to other computational studies of conversations. As in the mental health setting, these studies have largely considered the behavior of interlocutors in a single conversation, tying it to outcomes such as persuasion, problem-solving and incivility (Curhan and Pentland, 2007; Rosenthal and McKeown, 2015; Wang et al., 2017; Zhang et al., 2018), or to social correlates such as power, gender and age (Gonzales et al., 2010; Ireland et al., 2011; Danescu-Niculescu-Mizil et al., 2012; Nguyen et al., 2013; Prabhakaran et al., 2014). In contrast, our work seeks to chart the development of linguistic tendencies of individuals as they engage in many conversations.

**Linguistic change.** Prior studies have examined language change in offline (Labov, 1966, 1972, 2011; Eckert, 2005; Tagliamonte, 2011) and online (Cassell et al., 2006; Lam, 2008; Nguyen and Rosé, 2011; Garley and Hockenmaier, 2012; Danescu-Niculescu-Mizil et al., 2013; Bamman et al., 2014; Eisenstein et al., 2014; Nguyen et al., 2015; Kulkarni et al., 2015, 2016; Goel et al., 2016) settings. A core component of such studies of linguistic *socialization* has been the presence of external social influences that may impose linguistic norms—for instance, individuals are often able to observe the language used in posts by other members of an online community. In contrast, such social pressures are minimized in the counseling domain we consider. Our study hence reflects an alternate set of possible mechanisms for linguistic change, disentangling individual experience from normative influence.

### 3 Domain: Mental Health Counseling

In this work we study the linguistic development of counselors in Crisis Text Line, a large text-based mental health counseling platform. The platform offers a free 24/7 service for individuals undergoing mental health crises—henceforth *texters*—to have one-on-one conversations with

counselors via text messages. In collaboration with the platform, and in accordance with an IRB, we accessed the complete collection of anonymized conversations up to September 2018.<sup>2</sup>

**Counselors.** The service is driven by a dedicated roster of mental health counselors, comprised of volunteers who join the platform via an application process.<sup>3</sup> Each counselor starts by completing a standardized training curriculum; once they graduate they can then sign up for shifts during which they take conversations with texters. To gain a better understanding of the domain, the first author completed the training curriculum and participated in a shift.

Many counselors are quite committed to the service: the median counselor takes 43 conversations while a quarter take at least 120. In taking a longitudinal view, we focus on this latter subpopulation of counselors who have taken at least 120 conversations, and who started their tenure after July 2015.<sup>4</sup> The 3,475 counselors in this subset took their first 120 conversations over an average of 5 months; in total they have taken 1,055,924 conversations, accounting for 73% of all the conversations on the platform to date.

We analyze a counselor’s behavior at different experience levels by dividing their tenure into a series of consecutive *life-stages*  $S_0, S_1, \dots$  of  $s$  conversations each; unless otherwise indicated we take  $s = 20$ . Our particular focus is on examining counselor behavior over their first 120 conversations. We refer to a counselor as *tenured* once they have taken at least 100 conversations, and denote the life-stage consisting of a counselor’s 100th to 120th conversations as  $\bar{S}$ . Our subsequent analyses compare behavior in earlier life-stages to  $\bar{S}$ , using this eventual tenured state as a reference point.

**Counseling conversations.** In each conversation, a counselor is faced with a challenging task: guid-

---

<sup>2</sup>This study was done as part of a research fellowship program organized by Crisis Text Line, which granted the authors access to an anonymized version of the dataset, subject to an IRB. The program is open to other researchers by application: <https://www.crisistextline.org/open-data-collaborations>. Pisani et al. (2019) describes the extensive privacy and ethical considerations, and the policies implemented by the platform to address them.

<sup>3</sup>Just over half of counselors have no prior experience in a psychology-related domain. Our results are qualitatively similar over the subset of counselors who do not have a psychology-related background.

<sup>4</sup>We enforce the start time to account for variations in the training curriculum; our results are qualitatively similar under slight modifications to the start month, and over counselors with shorter tenures.

ing a texter through a moment of crisis towards a better mental state. This process requires the counselor to empathetically engage with the texter, often in difficult, high-stakes situations. The complexity of this inherently interactional task gives rise to substantive conversations—averaging 14 counselor messages per conversation and 28 words per counselor message—that yield a rich array of linguistic behaviors (Atkins et al., 2014; Althoff et al., 2016). We focus on tracking counselor behavior over conversations with at least 10 counselor messages (comprising 64% of our dataset).

The challenge of taking conversations is compounded by the diversity of texters who contact the service. Texters can come with a wide range of issues, from anxiety to family problems to suicidal ideation. The platform assigns each texter to an available counselor; under the assignment mechanism, counselors cannot select which conversations they take, such that longitudinal changes in behavior cannot be accounted for by shifts in counselors’ preferences. In addition, many texters only contact the service once and the mechanism does not specifically assign returning texters to the same counselor, distinguishing this setting from psychotherapy contexts with sustained counselor-client relationships considered in other work.

One salient guiding force for new counselors in this challenging domain is the training curriculum, through which counselors develop skills such as active listening and collaborative problem-solving in a 35-hour-long course. Beyond this initial training, other forces may also act to shape a counselor’s behavior through the remainder of their tenure. For instance, as counselors take more conversations, they are exposed to a broader range of texter situations beyond what the curriculum can account for; these experiences may further reinforce or alter their linguistic practices and fill inevitable gaps in the training.<sup>5</sup>

**Feasibility check: Is there linguistic change?** We first provide a coarse demonstration that the linguistic behavior of counselors does indeed change with experience in a systematic way. We consider a toy classification task: determining whether a conversation is taken by a *new* coun-

<sup>5</sup>Each shift is overseen by supervisors who can offer occasional feedback to counselors during their shifts; counselors can also interact with one another and solicit high-level advice. While in this context the level of social feedback is minimal, an interesting line of future work could more explicitly consider the interplay between an individual’s behavioral development and such interactions.

selor who has experienced less than 20 conversations ( $\mathbb{S}_0$ ), versus a *tenured* counselor who has taken between 100 and 120 ( $\bar{\mathbb{S}}$ ), on the basis of the language in the counselor’s messages. In particular, we formulate a *paired* prediction problem of distinguishing between two conversations taken by the *same* counselor at  $\mathbb{S}_0$  and  $\bar{\mathbb{S}}$ . Over a random subset of 10% of counselors in our data—comprising 3,075 conversation pairs—we train a logistic regression model using bigrams in counselor messages as features; we perform 10-fold cross-validation and ensure that no counselor spans multiple folds.<sup>6</sup>

The model gets a cross-validation accuracy of 86%, underlining that counselors undergo dramatic linguistic changes with experience; the high accuracy also highlights that such changes exhibit *systematic* consistencies across different counselors. Moreover, an analogous model that considers texters’ messages achieves only 57% accuracy, suggesting that changes in counselor behavior do not merely reflect changes in the texters they interact with.<sup>7</sup> This linguistic development is also not solely encompassed by population-wide shifts in trivial surface-level attributes—there is no systematic change in the word-counts of counselors’ messages or the lengths of their conversations.

In the remainder of this work, we will develop a methodology to characterize the systematic nature of this linguistic change.

## 4 Measures of Linguistic Diversity

As discussed in the introduction, certain circumstantial forces can drive counselors to diversify their conversational language with experience, while others can lead to linguistic stagnation. In order to operationalize these intuitions, we first design a general framework aimed at capturing the degree of *linguistic diversity* across conversations and individuals.<sup>8</sup> We then instantiate this framework in the counseling domain and esti-

<sup>6</sup>We use logistic regression models with  $\ell_1$  loss, tf-idf transforming features, and grid-searching over  $C$ , the number and the maximum document frequency of bigrams.

<sup>7</sup>That this low accuracy nonetheless exceeds the 50% random baseline suggests that changes in counselor behavior are reflected in aspects of a texter’s behavior as well. Indeed, prior work (Althoff et al., 2016) has suggested that texters can be influenced within a conversation; future work could further model whether the nature of this influence on the texter also varies with counselor experience.

<sup>8</sup>To encourage its application to other conversational domains, we release an implementation of this framework as part of ConvoKit: <https://convokit.cornell.edu>

mate the relation between the linguistic diversity of counselors and their effectiveness in engendering positive outcomes. Finally, we use the resulting high-level linguistic characterization to analyze the *evolution* of a counselor’s behavior over their tenure—the main focus of this work.

**Types of diversity.** We can measure linguistic diversity in conversations relative to two reference points. First, an individual can linguistically vary *across the different conversations* they have; second, their language can deviate from that of *other individuals*. We develop two complementary language-model-based measures corresponding to these reference points.

**Within-individual diversity.** Formally, for a particular individual  $\mathcal{I}$  and life-stage  $\mathbb{S}_i$ , we divide the set of conversations in that stage into two temporally-interleaved subsets, henceforth the *training* and *test* sets. We train a unigram language model  $\mathcal{L}_{\mathcal{I}}^i$  over all of the individual’s messages in conversations in the training set. For each conversation  $c$  in the test set, we can then compute its *within-cross-entropy*  $H(c, \mathcal{L}_{\mathcal{I}}^i)$  to  $\mathcal{I}$ ’s language model in  $\mathbb{S}_i$ . We average within-cross-entropies across all test set conversations to quantify  $\mathcal{I}$ ’s within-individual diversity during  $\mathbb{S}_i$ .<sup>9</sup>

**Between-individual diversity.** We can analogously define  $\mathcal{I}$ ’s between-individual diversity by replacing  $\mathcal{I}$ ’s language model at life-stage  $\mathbb{S}_i$  with the language model of another individual  $\mathcal{J}$  at the same life-stage, where  $\mathcal{J}$  is randomly-selected among  $\mathcal{I}$ ’s peers. In this way, we calculate the *between-cross-entropy*  $H(c, \mathcal{L}_{\mathcal{J}}^i)$  for each conversation  $c$  in  $\mathcal{I}$ ’s test set of conversations, which we average to quantify  $\mathcal{I}$ ’s between-individual diversity during  $\mathbb{S}_i$ .

Cross-entropy is known to be sensitive to the amount of text used to train  $\mathcal{L}_{\mathcal{I}}^i$ , as well as to the length of the document  $c$  (Genzel and Charniak, 2002). To mitigate trivial length-based effects and ensure that our measures of diversity afford meaningful comparisons across life-stages and across individuals, we randomly sample  $W$  words from each training set to construct the language models, and  $w$  words from each conversation to compute the cross-entropies. For robustness, we take averages over several samples for each measure, and over randomly selected peers  $\mathcal{J}$  for between-cross-entropy.

<sup>9</sup>For computational efficiency, we account for unseen vocabulary with an approximate form of smoothing by treating unseen words as tokens with frequency 1 in the training set.

**Relative diversity.** We note that within- and between-individual diversity are intrinsically related. If an individual’s own conversations are closer to each other than they are from conversations of other individuals we can say that the counselor has a *distinctive voice*. This would not be the case if the between-individual diversity would be entirely explained by their within-individual diversity. To capture this intuitive notion we can measure  $\mathcal{I}$ ’s linguistic distinctiveness during each  $\mathbb{S}_i$  as the average difference between the within- and between-cross-entropies of each of their conversations in the corresponding test set. We interpret high relative diversities as indicative of individuals who have a consistent voice that stands out when compared to others.

**Application to the counseling domain.** To apply these methods to analyze mental health counselors, we construct language models from  $W = 2,000$  words of each counselor and life-stage, compute cross-entropies with  $w = 200$  words per conversation and average over  $i = 50$  random samples.<sup>10</sup> To account for potential differences in peer groups that may have been trained with different curricula, for between-counselor diversity we consider as *peers* counselors who started taking conversations in the same month.

**Relation to counselor effectiveness.** Before we analyze the dynamics of linguistic diversity, we briefly consider its downstream implications for conversation outcome—to what extent is linguistic diversity positively reflective of a counselor’s skill? Higher diversity can signal the ability to hold more natural interactions and to better address specific texter situations, which according to literature might translate into positive counseling outcomes (Gaume et al., 2010; Atkins et al., 2014; Pérez-Rosas et al., 2018).

Determining the quality of a counseling conversation is a fundamental difficulty in this domain (Tracey et al., 2014). As one imperfect indicator of quality, we make use of texter responses to a survey given after each conversation, asking the texter whether or not they felt helped by the interaction. In our data, 26% of conversations received a rating, out of which 87% were rated as helpful. We consider the proportion of positive ratings a counselor received in a particular life-stage as an indication of their effectiveness.<sup>11</sup>

<sup>10</sup>These parameters are chosen to ensure sufficient data was used to characterize a counselor’s language.

<sup>11</sup>Here, we ignore conversations where no response was

To gauge the implications of a counselor's linguistic diversity, we compare the effectiveness of the most and least diverse counselors, in terms of both diversity measures. In the following discussion, we focus on experienced counselors, in the life-stage comprising their 80th to 120th conversations,<sup>12</sup> and compare the top and bottom third of counselors for each measure.

We find that both within- and between-individual diversities are positive signals of effectiveness—for instance, the average effectiveness of the most within-individual diverse counselors is 89% while the effectiveness of the least within-individual diverse counselors is 86% (Mann Whitney U test  $p < 0.001$ ).

We note that the observed correspondences between diversity and effectiveness have a potentially trivial explanation: a counselor may have an easier time varying their language across a series of texters with less complicated issues who may more readily give good ratings, independent of the counselor's actual skill. To discount this somewhat circular explanation, we can estimate the counselor's diversity in an earlier life-stage (40th to 80th conversations), and relate it with their effectiveness in the *subsequent* life-stage (80th to 120th conversations). The above differences still hold with statistical significance in this new setting, showing that diversity is non-trivially informative of a counselor's *future* effectiveness.<sup>13</sup>

## 5 Analyzing Counselor Diversification

We are now equipped to analyze how linguistic diversity develops as counselors gain experience.

**Do counselors diversify?** We first quantify the extent to which any development occurs. To this ends, we compute the percentage of counselors that increased their diversity from their initial conversations during  $\mathbb{S}_0$  to their tenured life-stage  $\bar{\mathbb{S}}$ . We find that, overall, counselors tend to diversify as they gain experience: 58% increase in within-individual diversity—indicating that they grow to vary their language to a greater extent between

---

received and consider only counselors in life-stages where they received at least four ratings.

<sup>12</sup>Our results are similar, with some loss of significance, for alternate experience levels.

<sup>13</sup>Developing a model that can estimate the future effectiveness of a counselor constitutes an interesting avenue for future work. We note that this task would be distinct from that of predicting the effectiveness of a presently-observed conversation, as was addressed in prior work (Atkins et al., 2014; Althoff et al., 2016; Pérez-Rosas et al., 2018).

different conversations they take—, and 73% increase in between-individual diversity—indicating that they become more linguistically distant from their similarly-experienced peers. Taking these measures together, we also find that most counselors increase in relative diversity (72%), becoming linguistically further away from their peers than from themselves. Overall, these trends suggest an increasing inclination for counselors to move away from an initial linguistic rigidity, eventually finding their own distinctive voice.<sup>14</sup>

**When do counselors diversify?** Is this observed diversification sustained over the counselors' tenures, or concentrated in an initial change? Answering this question is important for understanding, and potentially assisting, the developmental trajectory of mental health counselors (Tracey et al., 2014). As before, we consider as a reference the language of tenured counselors during  $\bar{\mathbb{S}}$ , and show in Figure 1 (top row) the percentage of counselors that increase in diversity between each of their earlier life-stages and  $\bar{\mathbb{S}}$ . We observe that both types of diversity increase over an extended span of counselors' tenures, but gradually level off to match the diversity of  $\bar{\mathbb{S}}$  (i.e., percentages approach 50%), suggesting that linguistic development plateaus with experience.

**What diversifies?** Conversations are not uniform—different aspects of the interaction may engender differing evolution dynamics. In this particular domain, conversations follow a well-defined structure that is taught as a central element of the training curriculum, proceeding through functionally distinctive segments—henceforth *components*—where counselors (1) build initial rapport with the texter (*hello*), (2) explore the texter's problems (*problem exploration*), (3) identify the texter's goals and past attempts to cope (*goal identification*), (4) collaboratively explore steps towards achieving those goals (*problem solving*), and finally (5) close the conversation (*goodbye*). As a rough proxy to capture these functional components, we divide a conversation into fifths, as in prior work (Althoff et al., 2016).<sup>15</sup> Characteristic words for

---

<sup>14</sup>We note that the time it takes for the counselors to reach their tenured life-stage is only mildly correlated to each diversity measure ( $\rho < 0.2$  for each).

<sup>15</sup>For this analysis, we only consider conversations with at least 20 counselor messages to ensure meaningful conversation segments. While we found that this coarse segmentation was satisfactory for our analyses, future work could consider more involved approaches to segmenting conversations.

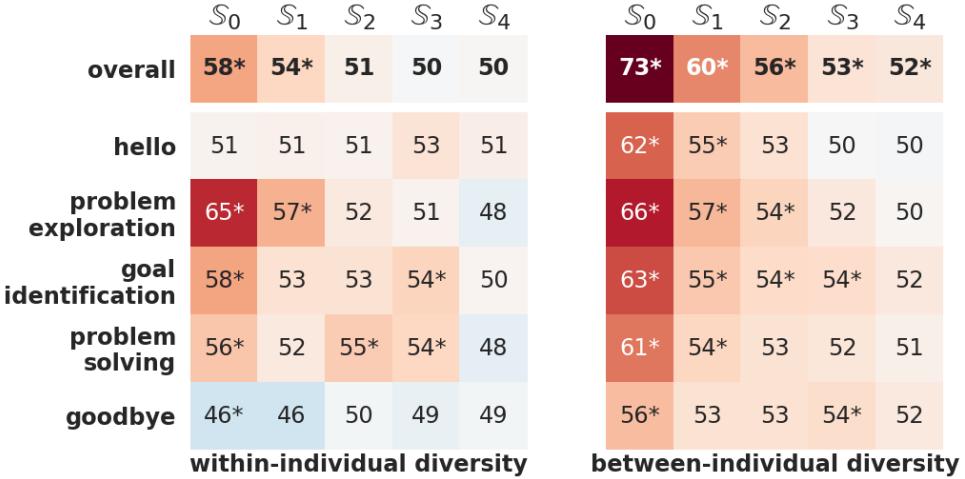


Figure 1: Temporal diversification trends across different counselor life-stages (left to right). Each cell shows the percentage of counselors that increase in diversity from that life-stage to their tenured life-stage  $\bar{S}$ ; \*'s indicate statistical significance (binomial test  $p < 0.05$ , comparing to 50% by chance). The topmost row shows the temporal trend across the entire conversation, and the subsequent rows separate this trend by conversational component.

each component are shown in Figure 2. To chart linguistic development with respect to our measures across functionally-different aspects of the conversation, we compute each measure separately over the messages in each fifth of the conversation, and repeat our temporal analysis per component (Figure 1, bottom rows).

We observe that development in within-individual diversity is largely concentrated in situation-specific segments of the conversation (problem exploration, goal identification, problem solving); in these segments the development is sustained throughout counselor tenures. This could suggest that as counselors encounter more diverse situations, they become linguistically better equipped to approach them in a specialized fashion. In contrast, in conversational components which have functions that generalize across texter problems there is no systematic increase in diversity (*hello*), and there is even a slight decrease (*goodbye*)—perhaps indicating that counselors develop routines to close the conversation.

On the other hand, between-individual diversity exhibits a comparatively more uniform increase across each component. This could suggest that counselors adopt individually distinguishing linguistic styles throughout the conversation.

In the next section we provide further support for these intuitions through a finer-grained analysis of vocabulary changes.

## 6 Development of Counselor Vocabulary

Thus far, we have tracked the evolution of counselors according to high-level characterizations of their linguistic diversification. We now analyze this evolution at a complementary granularity: tracking changes in counselors’ use of different *words* over the course of their tenure, as tangible and interpretable indicators of linguistic development. This analysis offers concrete examples of lexical changes that reflect the intuitions gleaned from our preceding examination.

A counselor may grow to use some words more often, perhaps reflecting continued exposure to particular texter situations or the development of a personal style. Such adoptions could enrich and diversify a counselor’s vocabulary, and perhaps distinguish them from others; they could also mark aspects that converge to a common counselor language, counterbalancing the diversification process. Alternatively, some words may fall out of favor, getting used less frequently over time. This possibility is especially pertinent given the initial training process, since the curriculum illustrates many of the counseling practices by way of presenting examples of messages for counselors to build on. These serve as a linguistic reference point for starting counselors; linguistic divergence could then arise as more experienced counselors start to move beyond this reference point.

To examine such changes in counselor vocabulary, we quantify the population-wide *usage shift* of a word as the log-ratio between the proportion

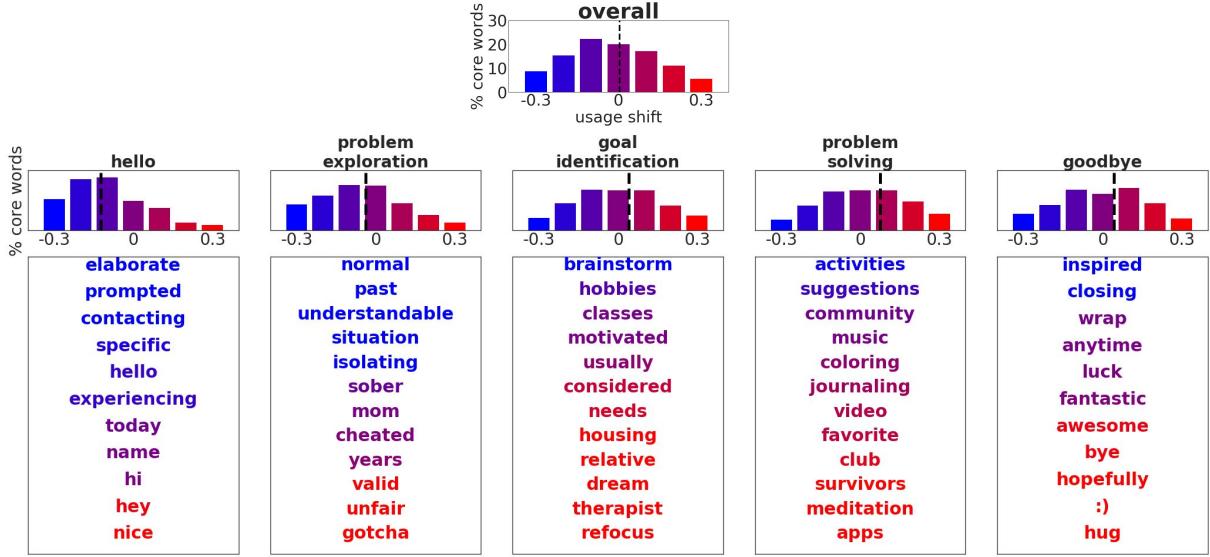


Figure 2: **Top:** the overall distribution of usage shifts of the words at the core of the counselor vocabulary (i.e., those used by at least 20% of counselors). **Bottom:** usage shift distributions per subset of core words characteristic to each conversational component, along with examples. Words are ordered by, and colored according to their usage shift, such that **blue** words tend to be used more by counselors in their earlier conversations while **red** words tend to be used more by counselors with more experience. Dashed lines indicate the respective medians.

of conversations in  $\bar{S}$  in which counselors used that word, and the proportion of conversations in  $S_0$  in which counselors used that word.<sup>16</sup> Thus, a positive shift indicates that the word’s usage increases with tenure, while a negative shift means that its usage declines after the counselor’s first conversations. Figure 2 (top) shows a histogram of usage shifts for the *core* of the counselors’ vocabulary, i.e., the 3,461 words used by at least 20% of the long-serving counselors considered in the preceding analyses.

**The interplay of lexical shifts and function.** Our prior analyses revealed that diversification patterns vary by functionally-different conversational components. To tie usage shifts to these observed contrasts, we examine the usage shifts of words characteristic to each component. Concretely, we consider a word to be characteristic of a component if the proportion of messages containing it in the corresponding fifth of the conversation is greater than the proportion of messages with the word overall (in particular, the log-ratio of the within-component frequency of the word to its overall frequency is at least 0.2). Figure 2 (bottom) shows usage shifts over words in each component, together with example words sorted by their shift.

<sup>16</sup>To avoid spurious effects akin to Simpson’s paradox, we only consider conversations taken by counselors with at least 120 conversations, such that each counselor contributes equally to both the  $\bar{S}$  and  $S_0$  components of the measure.

Across each component, we see that counselors shift away from using words which are relatively general and more formal in tone (negative shift, in blue). Many of these words reflect counseling practices as presented in the training material, echoing the language used in training to exemplify these practices. For instance, during the first component (*hello*), counselors are told to build rapport and bootstrap the exploration of the texter’s concerns by having texters *elaborate* on what *prompted* them to *contact* the platform; as counselors proceed in this exploration process (problem exploration) they are taught to reflect on and validate the texter’s concerns as *understandable* and *normal*.

That counselors eventually use such words less often does not necessarily mean that the practices underlying them are abandoned with experience; rather, counselors may gravitate towards new words that accomplish the same purposes (*valid*, *unfair*). Words used in the first two components are especially prone to decreasing in usage with experience. This suggests that while incoming counselors use the training material as a linguistic guide for initiating conversations, they tend to eventually shift towards more informal and less standardized language, moving away from using words in the core of the counselors’ vocabulary.<sup>17</sup>

<sup>17</sup>We note that our usage shift methodology only captures

Conversely, many words which counselors adopt later in their tenure (positive shift, in red) are more specialized, especially in the goal identification and problem solving components. Intuitively, counselors accumulate domain knowledge with experience, allowing them to better differentiate between cases, especially in terms of potential coping mechanisms (e.g., *therapist*, *meditation*, *apps*). This type of specialization suggests that much of the broadening in within-individual diversity we observed in these conversational segments (Figure 1, left) may be functional in nature.

Other words which increase in usage with experience are characterized by a relatively colloquial tone (*hey*, *:*, *gotcha*), pointing to a diversification process that is social rather than functional. This linguistic relaxation may trace the observed trend towards higher between-individual diversity (Figure 1, right) as counselors become more at ease in their task. Shifts towards such colloquial language are especially prominent in the last component (*goodbye*), potentially as counselors develop individual routines for closing conversations, and echoing the decrease in within-individual diversity observed in that component (Figure 1, left).

Overall, these observations complement our prior discussion about the rate at which counselors diversify their language, suggesting functional and social mechanics that may be in play.

## 7 Discussion and Future Work

Having good conversations is challenging and hard to teach, especially in domains where the stakes are high and where feedback is scarce. Computational tools could help us understand how people acquire conversational skills, and may eventually assist in their development. In this work we provide an initial and limited case study in a highly consequential domain, showing that we can track diversification in the linguistic practices of mental health counselors.

Our approach is necessarily limited in scope. Future work could adapt our framework to examine more complex forms of conversational development. In particular, we observed that counselors diversify linguistically, perhaps signaling a benefit-

---

population-wide changes in how counselors use a subset of core words, so it cannot characterize changes outside this general vocabulary. Future work could examine individual-level lexical changes (such as the development of uniquely personal catchphrases).

cial increase in flexibility that enables them to better address the specific concerns of a texter. Future work could more directly model how counselors respond to texter behaviors, hence gauging the extent to which counselors evolve in their *interactional* practices. Furthermore, while our approach captures linguistic changes in aggregate, a complementary line of work could explore the trajectories of individuals, and probe the factors determining whether particular individuals diversify or stagnate.

The framework we have started to design could eventually help platforms provide counselors with personalized feedback on their development. It could assist in identifying counselors who acclimate quickly and those that require more guidance. Understanding how different components in a conversation change with experience could also inform the particular aspects to focus their training on. In order to derive such prescriptive recommendations, further work is needed to causally connect our signals of linguistic experience with actual expertise and skill, as reflected in conversational outcomes. For instance, increased diversification may point to flexibility and specialization that are beneficial, but might also signal detrimental deviations from good counseling practices. Other approaches, such as qualitative labeling by domain experts, could examine whether such changes in language use also result in better conversations.

Our methodology could also be extended to examine other conversational contexts such as academic advising or business interactions, where individuals are expected to learn from experience. Such domains may contain crucial differences that motivate extensions to our framework; for example, feedback might be more readily available and conversational partners may recur, both of which can interact with experience to further shape linguistic development.

**Acknowledgments.** Thanks to Crisis Text Line for making this work possible. We are grateful to the anonymous reviewers for their thoughtful comments and suggestions, and to Jonathan P. Chang, Liye Fu, Sendhil Mullainathan and Andrew Wang for helpful conversations. This research has been supported in part by NSF CAREER Award IIS1750615, NSF Grant SES-1741441, and a Microsoft Research PhD Fellowship. The collaboration with Crisis Text Line was supported by the Robert Wood Johnson Foundation; the views expressed here do not necessarily reflect the views of the foundation.

## References

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics*, 4.

David C. Atkins, Mark Steyvers, Zac E. Imel, and Padhraic Smyth. 2014. Scaling up the evaluation of psychotherapy: Evaluating motivational interviewing fidelity via statistical text classification. *Implementation Science*, 9(1).

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18.

Justine Cassell, David Huffaker, Dona Tversky, and Kim Ferriman. 2006. The language of online leadership: Gender and youth engagement on the Internet. *Developmental Psychology*, 42.

Louis Georges Castonguay, Marvin R. Goldfried, Susan Wiser, Patrick J. Raue, and Adele M. Hayes. 1996. Predicting the effect of cognitive therapy for depression: A study of unique and common factors. *Journal of Consulting and Clinical Psychology*, 64(3).

Delwyn Catley, Kari Jo Harris, Matthew S. Mayo, Sandra Hall, Kolawole S. Okuyemi, Thuy Boardman, and Jasjit S. Ahluwalia. 2006. Adherence to Principles of Motivational Interviewing and Client Within-Session Behavior. *Behavioural and Cognitive Psychotherapy*, 34(1).

Munmun De Choudhury and Emre Kiciman. 2017. The Language of Social Support in Social Media and its Effect on Suicidal Ideation Risk. In *Proceedings of ICWSM*.

Jared R. Curhan and Alex Pentland. 2007. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Journal of Applied Psychology*, 92.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of WWW*.

Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities. In *Proceedings of WWW*.

Robyn Dawes. 1996. *House of Cards*. Free Press, New York.

Penelope Eckert. 2005. Variation, convention, and social meaning. In *Proceedings of the Annual Meeting of the Linguistic Society of America*.

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing. 2014. Diffusion of lexical change in social media. *PLoS One*, 9.

Sindhu Kiranmai Ernala, Asra F. Rizvi, Michael L. Birnbaum, John M. Kane, and Munmun De Choudhury. 2017. Linguistic Markers Indicating Therapeutic Outcomes of Social Media Disclosures of Schizophrenia. In *Proceeding of CSCW*.

Matt Garley and Julia Hockenmaier. 2012. Beefmoves: Dissemination, Diversity, and Dynamics of English Borrowings in a German Hip Hop Forum. In *Proceedings of ACL*.

Jacques Gaume, Nicolas Bertholet, Mohamed Faouzi, Gerhard Gmel, and Jean-Bernard Daepen. 2010. Counselor motivational interviewing skills and young adult change talk articulation during brief motivational interventions. *Journal of Substance Abuse Treatment*, 39(3).

Dmitriy Genzel and Eugene Charniak. 2002. Entropy Rate Constancy in Text. In *Proceedings of ACL*.

Rahul Goel, Sandeep Soni, Naman Goyal, John Parrizos, Hanna Wallach, Fernando Diaz, and Jacob Eisenstein. 2016. The Social Dynamics of Language Change in Online Networks. In *Proceedings of SocInfo*.

Simon B. Goldberg, Tony Rousmaniere, Scott D. Miller, Jason Whipple, Stevan Lars Nielsen, William T. Hoyt, and Bruce E. Wampold. 2016. Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology*, 63(1).

Amy L. Gonzales, Jeffrey T. Hancock, and James W. Pennebaker. 2010. Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37.

Ivan Habernal and Iryna Gurevych. 2016. What makes a convincing argument? Empirical analysis and detecting attributes of convincingness in Web argumentation. In *Proceedings of EMNLP*.

William P. Henry, Hans H. Strupp, Stephen F. Butler, Thomas E. Schacht, and Jeffrey L. Binder. 1993. Effects of training in time-limited dynamic psychotherapy: Changes in therapist behavior. *Journal of Consulting and Clinical Psychology*, 61(3).

Clara E. Hill, Ellen Baumann, Naama Shafran, Shudarshana Gupta, Ashley Morrison, Andrés E. Pérez Rojas, Patricia T. Spangler, Shauna Griffin, Laura Pappa, and Charles J. Gelso. 2015. Is training effective? A study of counseling psychology doctoral trainees in a psychodynamic/interpersonal training clinic. *Journal of Counseling Psychology*, 62(2).

Molly E. Ireland, Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker. 2011. Language style matching predicts relationship initiation and stability. *Psychological Science*, 22.

Daniel Kahneman and Gary Klein. 2009. Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6).

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of WWW*.

Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2016. Freshman or Fresher? Quantifying the Geographic Variation of Language in Online Social Media. In *Proceedings of ICWSM*.

William Labov. 1966. *The Social Stratification of English in New York City*. The Social Stratification of English in New York City. Washington, D.C.: Center for Applied Linguistics.

William Labov. 1972. *Sociolinguistic Patterns*. University of Pennsylvania Press.

William Labov. 2011. *Principles of Linguistic Change, Cognitive and Cultural Factors*. Wiley-Blackwell.

William Labov and David Fanshel. 1977. *Therapeutic Discourse: Psychotherapy as Conversation*. Academic Press.

Wan Shun Eva Lam. 2008. Language Socialization in Online Communities. In Nancy H. Hornberger, editor, *Encyclopedia of Language and Education*. Springer US.

William R. Miller and Kathy A. Mount. 2001. A small study of training in motivational interviewing: Does one workshop change clinician and client behavior? *Behavioural and Cognitive Psychotherapy*, 29(4).

Dong Nguyen, A. Seza Doğruöz, Carolyn P. Rosé, and Franciska de Jong. 2015. *Computational Sociolinguistics: A Survey*. *Computational Linguistics*.

Dong Nguyen, Rilana Gravel, Dolf Trieschnigg, and Theo Meder. 2013. How Old Do You Think I Am?: A Study of Language and Age in Twitter. In *Proceedings of ICWSM*.

Dong Nguyen and Carolyn P. Rosé. 2011. Language use as a reflection of socialization in online communities. In *Proceedings of the Workshop on Language in Social Media*.

Vlad Niculae and Cristian Danescu-Niculescu-Mizil. 2016. Conversational Markers of Constructive Discussions. In *Proceedings of NAACL*.

David E. Orlinsky and Michael Helge Rønnestad. 2005. *How Psychotherapists Develop: A Study of Therapeutic Work and Professional Growth*. How Psychotherapists Develop: A Study of Therapeutic Work and Professional Growth. American Psychological Association.

Grant Packard, Sarah G. Moore, and Brent McFerran. 2018. (I'm) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm Interactions. *Journal of Marketing Research*, 55(4).

Verónica Pérez-Rosas, Xuetong Sun, Christy Li, Yuchen Wang, Kenneth Resnicow, and Rada Mihalcea. 2018. Analyzing the Quality of Counseling Conversations: The Tell-Tale Signs of High-quality Counseling. In *Proceedings of LREC*.

Anthony R. Pisani, Nitya Kanuri, Bob Filbin, Carlos Gallo, Madelyn Gould, Lisa S. Lehmann, Robert Levine, John E. Marcotte, Brian Pascal, David Rousseau, Shairi Turner, Shirley Yen, and Megan L. Ranney. 2019. Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line. *Journal of Medical Internet Research*, 21(1).

Vinodkumar Prabhakaran, Emily E. Reid, and Owen Rambow. 2014. Gender and Power: How Gender and Gender Environment Affect Manifestations of Power. In *Proceedings of EMNLP*.

Yada Pruksachatkun, Sachin R. Pendse, and Amit Sharma. 2019. Moments of Change: Analyzing Peer-Based Cognitive Support in Online Mental Health Forums. In *Proceedings of CHI*.

Sara Rosenthal and Kathy McKeown. 2015. I Couldn't Agree More: The Role of Conversational Structure in Agreement and Disagreement Detection in Online Discussions. In *Proceedings of SIGDIAL*.

James Shanteau. 1992. *Running Head: Competence in Experts*.

Thomas M. Skovholt, Michael Helge Rønnestad, and Len Jennings. 1997. Searching for Expertise in Counseling, Psychotherapy, and Professional Psychology. *Educational Psychology Review*, 9(4).

Sali Tagliamonte. 2011. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Wiley-Blackwell.

Chenhao Tan, Vlad Niculae, Christian Danescu-Niculescu, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of WWW*.

Michael Tanana, Kevin A. Hallgren, Zac E. Imel, David C. Atkins, Padhraic Smyth, and Vivek Srikanth. 2015. Recursive Neural Networks for Coding Therapist and Patient Behavior in Motivational Interviewing. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology*.

Michael Tanana, Kevin A. Hallgren, Zac E. Imel, David C. Atkins, and Vivek Srikanth. 2016. A Comparison of Natural Language Processing Methods for Automated Coding of Motivational Interviewing. *Journal of Substance Abuse Treatment*, 65.

Terence J. G. Tracey, Bruce E. Wampold, James W. Lichtenberg, and Rodney K. Goodyear. 2014. Expertise in psychotherapy: An elusive goal? *The American Psychologist*, 69(3).

Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin. 2017. Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes. *Transactions of the Association for Computational Linguistics*, 5.

Diyi Yang, Robert Kraut, and John M. Levine. 2017. Commitment of Newcomers and Old-timers to Online Health Support Communities. In *Proceedings of CHI*.

Diyi Yang, Robert Kraut, Tenbroeck Smith, Elijah Mayfield, and Dan Jurafsky. 2019. Seekers, Providers, Welcomers, and Storytellers: Modeling Social Roles in Online Health Communities. In *Proceedings of CHI*.

Diyi Yang, Miaomiao Wen, and Carolyn Penstein Rosé. 2015. Weakly Supervised Role Identification in Teamwork Interactions. In *Proceedings of ACL*.

Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and Self-Harm Risk Assessment in Online Forums. In *Proceedings of EMNLP*.

Justine Zhang, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Nithum Thain, Yiqing Hua, and Dario Taraborelli. 2018. Conversations Gone Awry: Detecting Early Signs of Conversational Failure. In *Proceedings of ACL*.

Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil. 2016. Conversational flow in Oxford-style debates. In *Proceedings of NAACL*.