Patient Education and Counseling xxx (2019) xxx–xxx

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou

Story Arcs in Serious Illness: Natural Language Processing features of Palliative Care Conversations

Lindsay Ross[a], Christopher M. Danforth[b], Margaret J. Eppstein[c], Laurence A. Clarfeld[c], Brigitte N. Durieux[a], Cailin J. Gramling[a], Laura Hirsch[a], Donna M. Rizzo[d], Robert Gramling[e,*]

[a] University of Vermont, Burlington, VT, USA
[b] Department of Mathematics, University of Vermont, Burlington, VT, USA
[c] Department of Computer Science, University of Vermont, Burlington, VT, USA
[d] Department of Civil Engineering, University of Vermont, Burlington, VT, USA
[e] Department of Family Medicine, University of Vermont, Burlington, VT, USA

ARTICLE INFO

Article history:
Received 15 July 2019
Received in revised form 17 November 2019
Accepted 19 November 2019

Keywords:
Stories
Conversation
Communication
Machine Learning
Natural Language Processing
Palliative Care
Artificial Intelligence
Narrative Analysis

ABSTRACT

Objective: Serious illness conversations are complex clinical narratives that remain poorly understood. Natural Language Processing (NLP) offers new approaches for identifying hidden patterns within the lexicon of stories that may reveal insights about the taxonomy of serious illness conversations.
Methods: We analyzed verbatim transcripts from 354 consultations involving 231 patients and 45 palliative care clinicians from the Palliative Care Communication Research Initiative. We stratified each conversation into deciles of "narrative time" based on word counts. We used standard NLP analyses to examine the frequency and distribution of words and phrases indicating temporal reference, illness terminology, sentiment and modal verbs (indicating possibility/desirability).
Results: Temporal references shifted steadily from talking about the past to talking about the future over deciles of narrative time. Conversations progressed incrementally from "sadder" to "happier" lexicon; reduction in illness terminology accounted substantially for this pattern. We observed the following sequence in peak frequency over narrative time: symptom terms, treatment terms, prognosis terms and modal verbs indicating possibility.
Conclusions: NLP methods can identify narrative arcs in serious illness conversations.
Practice implications: Fully automating NLP methods will allow for efficient, large scale and real time measurement of serious illness conversations for research, education and system re-design.
© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Promoting high quality communication in serious illness is a national priority for 21st century healthcare. [1,2] Approximately 2.8 million people will die this year in the United States [3], accounting for more than 30 million physician visits during the last six months of life [4]. Despite this frequent use of healthcare, many people who are nearing the end of their life do not fully understand the expected trajectory of their illness or the treatment choices they face; will undergo diagnostic tests and procedures that they would not have wanted had they better understood their illness and treatment options; will suffer with otherwise controllable symptoms and spend the last months of their lives in ways they would not have wished.2]

Arriving at our current norms of poor serious illness communication in healthcare did not happen suddenly. The ways in which we select and train our learners; hire, reward and support our clinicians; build our buildings; establish our patient care workflows; document our outcomes; and pay for and market our services took time to evolve. We now face a massive and compelling task to re-design our healthcare system to communicate better with people who are seriously ill. Valid and systematic quality measurement is the fulcrum for re-designing healthcare. [5,6]

Serious illness conversation science faces two important barriers to systematic quality reporting. First, we lack sufficient empirical understanding about the features of naturally-occurring serious illness conversations, how those features coalesce in

* Corresponding author at: Rowell 235, Department of Family Medicine, University of Vermont College of Medicine, 106 Carrigan Drive, Burlington, VT, 05401 USA
E-mail address: robert.gramling@uvm.edu (R. Gramling).

https://doi.org/10.1016/j.pec.2019.11.021
0738-3991/© 2019 Elsevier B.V. All rights reserved.

observable patterns, and how the context in which those conversations happen influences the patterns that indicate high quality. Second, traditional human coding methods for measuring serious conversations are slow and resource intensive, thus precluding timely quality reporting in routine healthcare settings. [7] Rapid advances in Natural Language Processing (NLP) and artificial intelligence, however, are improving our capacity to recognize and interpret complex features of clinical conversations automatically [8–11]. Most notably, Sen *et al.* applied NLP methods to examine physician-patient conversations in outpatient oncology. [12] They observed that the frequency and patterns of words spoken by the physician could distinguish conversations subsequently rated by patients as higher *versus* lower quality communication. [12] In this work, we use NLP to analyze features of serious illness conversations in a setting typically characterized by high quality communication that, to our knowledge, has yet to be characterized using NLP methods: inpatient palliative care consultation [13–17].

Palliative care serious illness conversations center around concepts of suffering and shared decision making, both of which require acute understanding of the person who is ill. [18,19] Understanding the person – how they define who they are, how they make meaning from their experiences, what suffering means for them and how decisions might affect them– happens over an arc of conversation and frequently organizes in the form of narrative [20–22]. This work is conceptually grounded in a Narrative Analysis [23],[24] model of spoken stories proposed by the sociolinguist William Labov and endorsed in computational linguistics for the automated analysis of stories. [25] Labov proposes that understanding the meaning of spoken narrative requires close attention to the order in which the unfolding story is oriented in time, which topics are shared earlier and which later, and how the teller evaluates those topics [24,26].

Serious illness conversation narratives are complex, relational and dynamic. At the individual level, each conversation is unique. However, related work observes that computational methods can identify lexical patterns within the trajectory of unique narratives when they are analyzed together in very large samples. [27] Revealing these patterns helped literary science to better understand the taxonomy of modern fiction [28]. This study takes an initial step toward similar advances in healthcare communication science by describing the aggregate structure of serious illness conversation story arcs in the natural palliative care consultation setting.

2. Methods

2.1. Overview

This is a cross-sectional analysis of 354 verbatim transcripts of audio-recorded inpatient palliative care consultations. We programmed NLP methods to do three things. First, we divided each conversation into ten even segments of narrative time defined by the number of words in that conversation's transcript. Second, we identified the frequency and timing of specific target words or phrases representing a reference to time, illness, sentiment or possibility/desirability. Third, we described the trajectories of these words/phrase types across deciles of narrative time.

2.2. Context & Participants

As described more fully elsewhere, [29] the Palliative Care Communication Research Initiative (PCCRI) is a multisite observational cohort study, conducted between January 2014 and May 2016, at the University of Rochester (New York) and University of California, San Francisco. All hospitalized patients referred for palliative care consultation during the study period were eligible for the study if they met the following criteria: diagnosed with metastatic cancer; English-speaking; older than 21 years of age; able to consent for research or had an established healthcare proxy who was able to do so. Patients were excluded if they had a “Comfort Measures Only” designation on their Medical Orders for Life Sustaining Treatment form or were already receiving hospice care at the time of referral. The PCCRI enrolled 45 palliative care physicians, physician fellows or nurse practitioners and 240 patient participants. Four of these patients withdrew, three died, and two were discharged before the palliative care consultation happened. This analysis includes the 354 conversations among the 231 patients with at least one visit.

2.3. Data Collection

All participants completed informed consent and a brief self-report questionnaire at the time of enrollment (*i.e.,* same day as the palliative care consultation). Prior to entry of the palliative care team, a research assistant unobtrusively placed a small digital audio-recorder with a built-in multidirectional microphone in the hospital room and initiated the recording. After the consultation, the research coordinator ended the recording. All participants were instructed how to stop recording should they wish; none did so.

Up to the first three visits with the palliative care team were audio-recorded and subsequently transcribed verbatim. Verbatim transcripts were formatted for natural language processing without removal or re-interpretation of any speech content.

2.4. Natural Language Processing Measures

2.4.1. Temporal Reference

The Natural Language Toolkit (NLTK; www.nltk.org) is an open source package used for working with human language data, utilizing text processing libraries for part of speech classification. NLTK assigns a part of speech label for individual verbs, but does not consider verb phrases, presenting a problem for accurate assignment of temporal reference categories. Some commonly-used verb forms in the English language—such as the base form and the participle forms—cannot be categorized validly into their temporal referent categories without considering the preceding words of the verb phrase (*e.g.*, “will run”/“had run” or “was running”/”am running”/“will be running”). Others, such as the present perfect progressive (*e.g.*, "has been running"), cross temporal categories. Therefore, we created a new NLP verb phrase dictionary based on the verbs in our serious illness conversation corpus to extend the existing NLTK tense assignment for verb forms requiring more lexical context. For verb phrases, especially those crossing temporal categories, each word of the phrase is labeled. For example, our NLP algorithm classifies "have been running" as two words labeled "present" (*i.e.*, "have", "running") and one as past (*i.e.*, "been"). This approach systematically and substantially reduces overall NLP misclassification. We anticipate that emerging ML methods may help to more fully resolve the complexity of inferring temporal reference in conversation. Our temporal reference NLP method is open source and available at www.vermontconversationlab.com.

2.4.2. Sentiment Score

We assess the sentiment of a palliative care conversation using an approximately 10,000 word sentiment dictionary called labMT (language assessment by Mechanical Turk). [30] Described in detail at hedonometer.org, this word list was developed by first combining the 5,000 most frequently used words found in each of four separate sources: three years of tweets, 20 years of *New York Times* articles, 200 years of books scanned by *Google*, and 60 years of music lyrics. In sum, a total of roughly 10,000 unique words were identified when combining the four sources. A total of 50 individuals then scored the sentiment of each of these words on a scale from 1 (sad), 5 (neutral), to 9 (happy). For example, the words “worse”, “of”, and “happy” received average scores of 2.77, 4.94, and 8.30 respectively. We calculated a sentiment score for each decile of narrative time by averaging word-related sentiment weighted for word frequency. As indicated during construction of labMT, we excluded words in the neutral range (4 to 6) in order to focus on words having a tangible impact on emotion during crowd-sourcing calibration. [30]

Since the labMT list was compiled and scored for sentiment in 2011, several studies have compared word scores across different corpora, participants, and languages. [31] While the health status of participants in those studies was not queried, labMT sentiment scores do correlate very strongly with several other dictionaries [32], and Twitter sentiment measured using labMT has been shown to correlate well with traditional survey-based measures of population level well-being [33].

2.4.3. Illness Terms

We created groupings of symptom, treatment, and prognosis terms to investigate how the usage of these terms fluctuated over time in palliative care conversations. To create groupings with relative stability, we only considered words used more than 100 times in the full conversation dataset. Among the 17,041 unique words in our dataset, 947 occurred more than one hundred times. We identified terms typically used when referring to symptoms, treatments or prognoses in this list of 947 words (see Appendix for final word lists).

2.4.4. Modal Verbs

Modal verbs are auxiliary verbs utilized to express desirability or possibility; they are used to show whether or not we believe something is certain, probable or possible. The modal verbs we identified for this analysis are “can”, “could”, “may”, “might”, “shall”, “should”, “will”, “would”, and “must”.

2.4.5. Human Subjects

The PCCRI study was approved by the Protection of Human Subjects committees at the University of Rochester, University of San Francisco and the University of Vermont.

2.5. Analytic Approach

2.5.1. Narrative Time

When analyzing trends in word usage over time, we use the concept of "narrative time", [27] where each conversation’s timeline is normalized to represent the percentage of the conversation words that have occurred up to that moment. For example, the appearance of the 100th word in a conversation of length 10,000 words would indicate 1% of narrative time has passed, whereas in a conversation of length 1,000 words it would indicate that 10% of narrative time has passed. To analyze trends in word use, we split the conversation into deciles of narrative time. We chose deciles to offer a sufficient number of categories with which to observe trends over narrative time while maintaining adequate numbers of observations per category for stable estimation. We conducted two sets of sensitivity analyses. First, we evaluated how categorizing narrative time into fewer or more groups (*e.g.*, 5, 25, 50, 100) affected our observed trends. Second, we excluded 50 conversations of fewer than 1,250 words (~10 minutes or shorter) or six outlier conversations longer than 10,000 words. Neither changed the interpretations of our findings.

2.5.2. Relative word frequency by decile of Narrative Time

Using the aggregate data of all conversations in the dataset, we calculated the number of times a type of word or phrase of interest appeared in each decile of narrative time. We then divided this frequency by the number of times the word or phrase of interest appears in the full conversation narrative. This creates a relative frequency by decile such that the sum of all decile relative frequencies will equal 1 for each word or phrase type. We used this relative frequency approach instead of an absolute frequency approach in order to directly visualize trajectories of words having different absolute frequencies in the same graphical image.

2.5.3. Estimating reliability of sentiment scores

To assess the reliability of the sentiment assigned to each decile, we calculated the sentiment of each decile 100 times with a random 10% of the words removed. This helps quantify the extent to which usage of any specific words substantially impacts a decile’s sentiment score, and ultimately demonstrates the statistical strength of the change in sentiment over time. In all cases, the observed trends remained qualitatively unchanged.

2.5.4. Sentiment Attribution by Word-Shift Graphs

In order to reveal the words responsible for differences in sentiment scores across deciles of narrative time, we use word-shift graphs as detailed in Dodds et al. [30] The sentiment of a reference group of words (*e.g.*, the average sentiment of all words that appeared in decile two) is measured against a comparison group of words (*e.g.*, the average sentiment of all words that appeared in decile nine). Words are rank-ordered by their contribution to the difference in sentiment between these two corpora and graphically represented in a histogram.

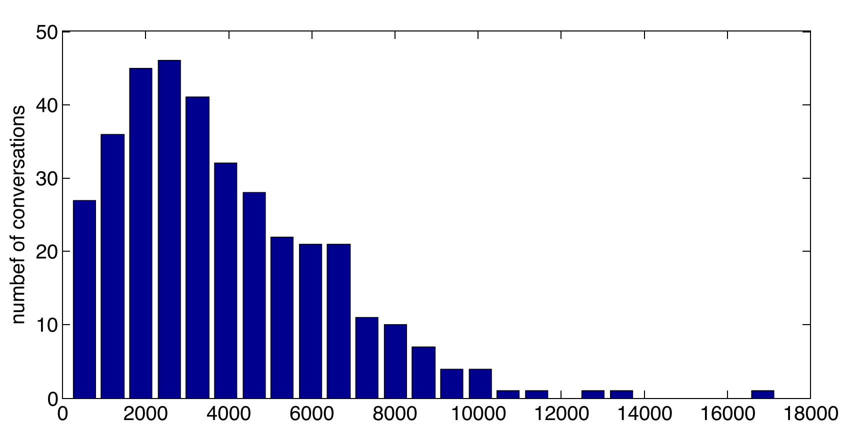

Fig. 1. Distribution of the number of words per conversation.

3. Results

Among the 231 patient participants, 114 (49%) were women, 29 (13%) identified as Black or African American, 62 (27%) were younger than age 55 years and 64 (28%) were older than 70 years, 140 (61%) were financially insecure and 67 (29%) completed a 4-year college degree. Metastatic cancers of the colon, breast or prostate were the most common (50, 22%), followed by those of the lung (49, 21%) and non-colon gastrointestinal tract (42, 18%). Patient participants lived for a median of 37 days (Interquartile range: 12 days, 97 days).

Among the 45 participating palliative care specialists, 25 (56%) were women and 16 (36%) were in palliative care practice for at least 5 years. Twenty-two (49%) were attending physicians, 13 (29%) were physician fellows, and 6 (13%) were nurse practitioners.

The number of words per conversation demonstrated a skewed distribution, ranging from 196 to 17,200 with a median of 3,355 words per conversation (Fig. 1).

3.1. Temporal Reference

Verbs/verb phrases referring to the present are roughly 3 times more prevalent than references to either past or future but show no significant change in frequency of usage over the course of the conversation (Fig. 2a). However, references to the past decrease and references to the future increase over narrative time (Fig. 2a) such that the relative proportions of temporal reference to the future compared to the past demonstrate a nearly monotonic increase as the shared narrative unfolds (Fig. 2b).

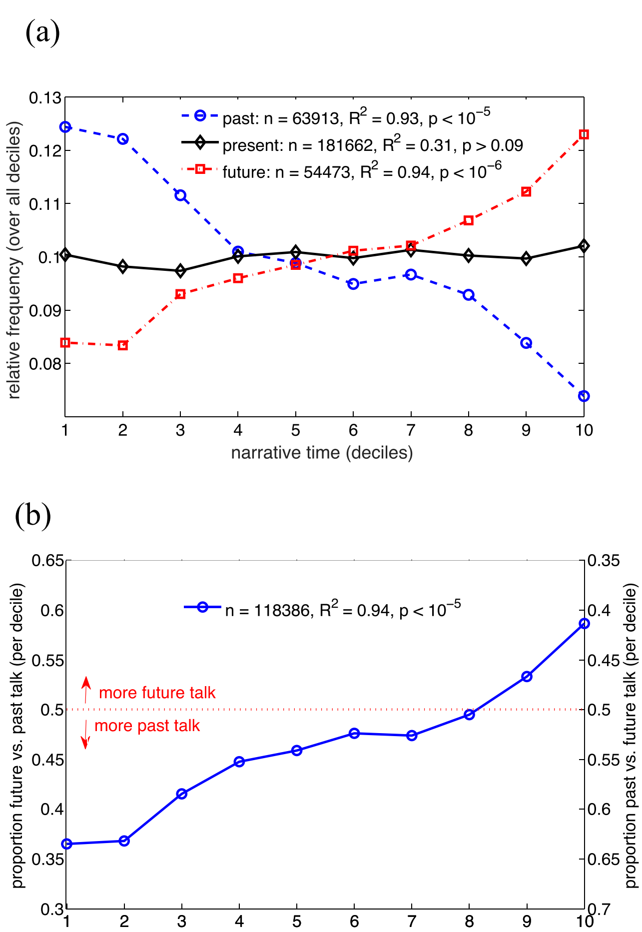

Fig. 2. Change in temporal reference over narrative time.
Legend: Change in temporal reference over narrative time. Fig. 2a) Distribution of past, present and future references across deciles; each curve is normalized across all deciles, such that the data points on each curve sum to 1.0. Fig. 2b) Relative proportion of past and future references, such that the proportion of future (left y-axis) is equal to 1 – the proportion of past (right y-axis) within each decile. The legends show the total number of words/phrases per curve and the statistics from linear regression; the regression lines are not shown to avoid clutter.

3.2. Sentiment Scores

Patterns in sentiment of the lexicon suggest that conversations become “happier” with the passage of narrative time (Fig. 3). The sentiment score is 5.91 in the first decile, drops to 5.82 in the second decile, then increases in a stepwise fashion to reach 6.08 in the final decile. To put these sentiment scores into perspective, we made comparisons to the Hedonometer [30] website (http://hedonometer.org), which scores a random 10% of all English tweets daily using the same sentiment method in this study. On days when tragic world events happened, the scores were similar to deciles one and two (*e.g.*, 2017 terrorist attack in Barcelona = 5.92; 2016 mass shooting in Pulse nightclub in Orlando = 5.84). In contrast, some common U.S. holidays exhibit scores similar to decile ten (*e.g.*, Easter 2017 = 6.08; Mother’s Day 2018 = 6.09). The change in sentiment over time in palliative care conversations thus moves from “sad” to “happy” in terms of societally assigned word valences.

The first and last deciles demonstrate more frequent use of higher sentiment score words and less frequent use of lower sentiment score words than in their nearest decile neighbor. Compared to decile two, the first decile includes greeting words (*i.e.*, “hi” and “hello”) and terms such as “okay”, “good”, and “nice” that have higher sentiment score values more often and uses negation words such as “not”, “don’t”, “doesn’t”, “didn’t”, and “no” that have lower sentiment score values less often. Similarly, in comparison to the ninth decile, the last decile includes more gratitude or departing words (*e.g.*, “thank”, “thanks”) that have high sentiment values more often and negation words less often. Both the first and the last decile use illness terms (*e.g.*, “cancer”, “pain” and “death”) substantially less frequently than their neighboring decile.

When excluding the beginning (*i.e.*, greeting) and ending (*i.e.*, departing) deciles of the conversation narrative, decreasing use of illness words (*e.g.*, "cancer", "pain") contributed substantively to the rise in sentiment (Fig. 4).

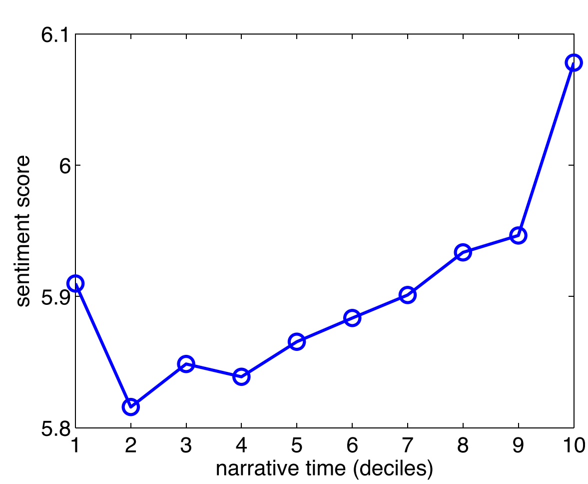

Fig. 3. Change in sentiment score over narrative time.

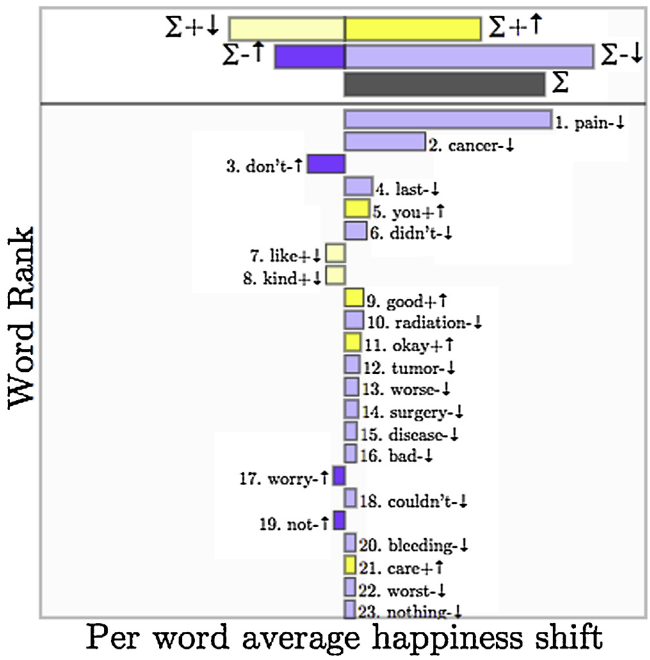

Fig. 4. Word shift histogram of top words affecting sentiment score between deciles two and nine.
Legend: The top 23 words driving the change in sentiment score from the second to the ninth decile. The right side represents which word differences contribute to increasing the sentiment score and the left represents which word differences contribute to lowering the sentiment score. The + and – symbols and yellow and blue colors indicate direction of the assigned word valence (i.e., + (yellow) means happier than neutral; - (blue) means sadder than neutral). The up/down arrows indicate whether a word was used more or less frequently. The summary bars at the top indicate that a decrease in sadder terms is the largest contributor to the increase in sentiment score from decile 2 to decile 9.

3.3. Illness Terms and Modal Verbs

The frequencies of symptom, treatment, and prognosis terms peak successively during deciles 2, 4, and 6, respectively and drop substantially during the final third of the conversation (Fig. 5). The relative frequency of modal verbs increases fairly steadily over narrative time, peaking in decile 9 (Fig. 5).

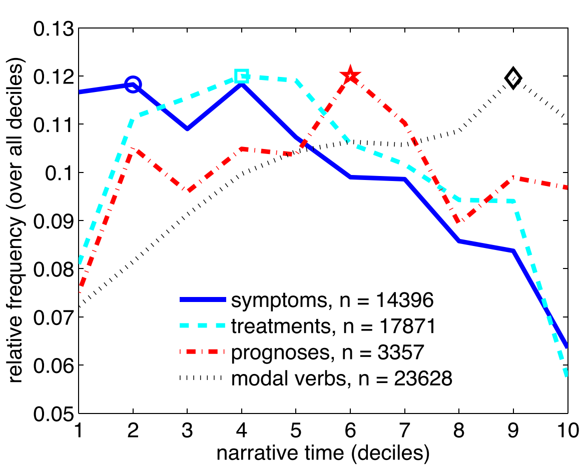

Fig. 5. Illness Terms and Modal Verbs.
Legend: Changes in illness terms and modal verbs over narrative time. Each curve is normalized across all deciles, such that the data points on each curve sum to 1.0. The symbols indicate the deciles with the largest magnitudes for each curve and the legend indicates the absolute number of terms in each curve.

3.4. Modal Verbs

We see that the relative frequency of modal verbs increases over narrative time, peaking in the ninth decile (Fig. 5).

4. Discussion and Conclusions

4.1. Discussion

We used NLP to evaluate the lexicon of naturally-occurring palliative care conversations and observed four patterns in word usage over narrative time that present potential feature targets for further investigation. First, we observed that participants' reference to time changed substantially over the course of a conversation. The conversation narrative progressed steadily from referencing the past more frequently than the future at the beginning to the future more frequently than the past at the end. Second, the sentiment associated with the words used from the beginning to the end of the conversation progressed steadily from "sadder" to "happier". We observed that this change was due, in large part, to changing patterns in use of illness terminology. Third, the trajectories of terminology referring to symptoms, treatment and prognosis peaked sequentially over time. Last, the use of modal verbs indicating possibility/desirability rose over the course of the conversation, peaking after prognosis terms. Below, we organize our thoughts about what hypotheses these findings might catalyze regarding our understanding of palliative care conversations and how these feature targets might be valuable in forthcoming research.

How speakers reference time and what this means in human discourse have long been foci for linguists and philosophers of language. [34,35] The purpose of Palliative Care is to understand, prevent and treat suffering. Human suffering is a highly variable existential experience that can have its roots in the past, present, and future [36]. Therapeutic conversation about suffering is dynamic and relational, often moving fluidly between painful meanings and joyful ones. The patterns of verb-related temporal reference exhibited in our findings suggest that time is an important dimension of the unfolding narrative in palliative care conversations. Often times, understanding whether an English speaking person is talking about the past, present or future requires more context than the verb conjugation or verb phrase (*e.g.*, "I am at a conference next week."). [34,35,37] Our analysis of temporal reference advances previous NLP methods by expanding interpretation of single base verbs or verb participles (*e.g.*, "laugh", "laughing") to verb phrases (*e.g.*, "will laugh"; "was laughing") and improving the categorization of temporal reference. Our method will still make errors given the complexity of the ways the people use language in conversation. For example, the algorithm will occasionally mistake nouns and verbs (*e.g.*, "laugh") and categorize some future and past references as "present" but we do not expect that this substantially changes the observed global trajectory of the narrative progressing from "past" to "future".

Second, we observed that the emotional valence of the words used in palliative care conversations rises incrementally towards a more positive global sentiment over narrative time. The PCCRI cohort study focused on clinical contexts where specialty palliative care was asked to visit with acutely ill hospitalized patients and their families specifically to help with decision-making about medical treatments in the context of advanced cancer. [29] These are often emotion-filled conversations in which patients contemplate medical prognoses and available treatment options amid the terror of potential death and dying [13,14,16,38–40]. In fact, as described in the Results, the sentiment score of words used in the early part of the conversation align with terrifying events. When people are experiencing intense emotions such as fear of dying, the ability to consider new information and reason effectively is quite challenging [41]. The observed trajectory of word-related sentiment during the conversation narrative might be a marker of the therapeutic process of palliative care conversations to make some emotional space to reason. We observed that patterns in illnessterms accounted for much of this shift in sentiment. The approach that we used here to assign a sentiment value to each word was developed by state-of-the-art crowd sourcing. The "crowd" was not made up of hospitalized people with advanced cancer. How this affects our findings is unclear. On the one hand, terms like "cancer", "blood" or "death" might have lower impact on sentiment for people who might have acculturated to hearing them in their own healthcare, thus biasing our findings. On the other hand, palliative care conversations typically happen amid confusion, terror and vulnerability. In these cognitive states, people often interpret meaning heuristically. Therefore, it is also quite possible that the crowd sourced sentiments assigned to words might *underestimate* their impact on emotions when hearing them during palliative care conversations. Validly assigning meanings to words is critically important for using NLP to measure clinical communication. More research is necessary to understand the strengths and weaknesses of crowd sourcing methods and to identify which "crowds" are appropriate for improving NLP of serious illness conversations.

This study has important limitations. First, these conversations include English speaking participants from two regions of the United States. It is likely that the linguistic characteristics of conversations among non-English speakers and those from different geographic regions will differ. Whether those potential differences would lead to different trajectories of sentiment, temporal reference, illness terms or modal verbs is unknown. Second, as mentioned above, we used word-associated sentiment scores as defined by crowd-sourcing among people who were not seriously ill. Crowd sourcing to attribute meaning to language-in-use is an important method for NLP, including for sentiment analyses of oncology conversations. [12] For serious illness communication research, such crowd-sourcing methods will benefit from including people with advanced diseases. Third, as we describe above, our existing NLP algorithm for temporal reference might underestimate the frequency of references to the past or future. Further work will require classification of non-verb lexicon to more accurately contextualize temporal reference in settings when the verb phrase is insufficient. Fourth, we evaluated usual trajectories of words in our sample of palliative care conversations. We anticipate, however, that palliative care conversations will exhibit fundamental sub-types that organize into a clinically relevant taxonomy of serious illness conversations. Further work is necessary to evaluate whether the features we identify in this analysis cluster into different trajectories that reveal distinct story arcs to palliative care conversations.

4.2. Conclusions

NLP is a useful method for empirically understanding the narrative arc of palliative care conversations. Our findings suggest that palliative care serious illness conversations are oriented from the past toward the future, from topics of symptoms to treatments to prognoses, from fewer to more indicators of possibility and desirability, and from sadder toward happier lexicon. This computational narrative approach holds promise for developing a taxonomy of serious illness conversations. Future research is needed to confirm these findings and establish whether these and other feature targets coalesce into identifiable story arcs that define clinically important sub-types of conversations.

4.3. PRACTICE IMPLICATIONS

Ecological theories of communication endorse a complex relationship between the context (habitat) and characteristics of conversations (phenotypes) that are beneficial for seriously ill persons' decision-making and amelioration of suffering. Our findings suggest that computational narrative analysis can reveal insights about the phenotpes of serious illness conversations that happen in the natural palliative care setting. Fully automating these computational methods for real time classification of serious illness conversation types will catalyze the capacity for observational researchers to conduct efficient large-scale studies to understand context-type (eg. habitat-phenotype) interactions, clinical trial researchers to develop type-matched interventions, educators to systematically evaluate their trainees' communication learning environment, and, eventually, healthcare re-design scientists to implement measurement-feedback systems that foster clinical environments where all seriously ill people experience meaningful, compassionate and person-centered communication.

Acknowledgments

This work was funded in part by a Research Scholar Grant from the American Cancer Society (RSG PCSM124655; PI: Robert Gramling), a National Science Foundation Award (S-STEM, DUE-1356426, PI: Margaret J. Eppstein), a unrestricted philanthropic gift from MassMutual Insurance to the Vermont Complex Systems Center (Danforth), a Vermont EPSCoR grant from the National Science Foundation (Rizzo), a Research Experience for Undergraduates Award from the University of Vermont College of Engineering and Mathematical Sciences, and the Holly & Bob Miller Endowed Chair in Palliative Medicine. We thank the palliative care clinicians, patients, and families who participated in this research for their dedication to enhancing care for people with serious illness.

Appendix A

Symptom, Treatment and Prognosis Terms identified among 947 words occurring at least one hundred times in the PCCRI dataset

Symptom

comfortable, worried, tired, painful, symptom, shortness, hurting, confused, uncomfortable, weak, happy, comfort, sleepy, depressed, hurts, symptoms, pain, breathing, cough, constipation, dry, energy, appetite, awake, hurt, coughing, sleep, breathe, strength, breath, sleeping, bothering, nausea, strong, anxiety, wake, scary, depression, worry, stronger, anxious

Treatment

morphine, patch, medications, drug, trial, CPR, line, Tylenol, button, doses, drugs, medical, feeding, oxygen, Ativan, Oxycodone, therapy, Dilaudid, chemotherapy, machine, antibiotics, treatment, radiation, surgery, treat, dose, meds, medicines, fluids, tube, hospice, medicine, dialysis, methadone, oral, ventilator, milligrams, management, resuscitation, fentanyl, chemo, pill, nutrition, ICU, milligram, medication, procedure, liquid, treatments, IV, pills

Prognosis

cure, future, dying, die, prognosis, probably, hope, risk, hoping, death

References

[1] National Priorities and Goals, Aligning Our Efforts to Transform America's Healthcare, Paper presented at: National Quality Forum, Washington, D.C, 2008.
[2] Institute of Medicine, Dying in America:Improving Quality and Honoring Individual Preferences Near End-of-Life, The National Academies Press, Washington, D.C, 2014.
[3] Center for Disease Control and Prevention, National Center for Health Statistics, (2019) , pp. 2019. https://www.cdc.gov/nchs/fastats/deaths.htm.
[4] Dartmouth Atlas of Healthcare: End-of-Life Care, (2017) . http://www.dartmouthatlas.org/keyissues/issue.aspx?con=2944.
[5] D.M. Berwick, B. James, M.J. Coye, Connections between quality measurement and improvement, Med Care. 41 (1 Suppl) (2003) I30–38.
[6] Institute for Healthcare Improvement, Science of Improvement, (2019) . http://www.ihi.org/about/Pages/ScienceofImprovement.aspx.
[7] J.A. Tulsky, M.C. Beach, P.N. Butow, et al., A Research Agenda for Communication Between Health Care Professionals and Patients Living With Serious Illness, JAMA Intern Med. (2017).
[8] P. Ryan, S. Luz, P. Albert, C. Vogel, C. Normand, G. Elwyn, Using artificial intelligence to assess clinicians' communication skills, BMJ. 364 (2019) l161.
[9] M. Tanana, K.A. Hallgren, Z.E. Imel, D.C. Atkins, V. Srikumar, A Comparison of Natural Language Processing Methods for Automated Coding of Motivational Interviewing, J Subst Abuse Treat. 65 (2016) 43–50.
[10] D. Can, R.A. Marin, P.G. Georgiou, Z.E. Imel, D.C. Atkins, S.S. Narayanan, It sounds like . . . ": A natural language processing approach to detecting counselor reflections in motivational interviewing, J Couns Psychol. 63 (3) (2016) 343–350.
[11] J.P. Pestian, J. Grupp-Phelan, K. Bretonnel Cohen, et al., A Controlled Trial Using Natural Language Processing to Examine the Language of Suicidal Adolescents in the Emergency Department, Suicide Life Threat Behav. 46 (2) (2016) 154–159.
[12] T. Sen, M.R. Ali, M.E. Hoque, R.M. Epstein, P. Duberstein, Modeling Doctor-Patient Communication with Affective Text Analysis, Paper presented at: Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, 2017.
[13] L.T. Ingersoll, F. Saeed, S. Ladwig, et al., Feeling Heard and Understood in the Hospital Environment: Benchmarking Communication Quality Among Patients With Advanced Cancer Before and After Palliative Care Consultation, J Pain Symptom Manage. 56 (2) (2018) 239–244.
[14] R. Gramling, M. Sanders, S. Ladwig, S.A. Norton, R. Epstein, S.C. Alexander, Goal Communication in Palliative Care Decision-Making Consultations, J Pain Symptom Manage. 50 (5) (2015) 701–706.
[15] S.C. Alexander, S. Ladwig, S.A. Norton, et al., Emotional distress and compassionate responses in palliative care decision-making consultations, J Palliat Med. 17 (5) (2014) 579–584.
[16] S.A. Norton, M. Metzger, J. DeLuca, S.C. Alexander, T.E. Quill, R. Gramling, Palliative care communication: linking patients' prognoses, values, and goals of care, Res Nurs Health. 36 (6) (2013) 582–590.
[17] R. Gramling, S.A. Norton, S. Ladwig, et al., Direct observation of prognosis communication in palliative care: a descriptive study, J Pain Symptom Manage. 45 (2) (2013) 202–212.
[18] E.J. Cassel, The nature of suffering and the goals of medicine, N Engl J Med. 306 (11) (1982) 639–645.
[19] R. Gramling, S. Stanek, S. Ladwig, et al., Feeling Heard and Understood: A Patient-Reported Quality Measure for the Inpatient Palliative Care Setting, J Pain Symptom Manage. 51 (2) (2016) 150–154.
[20] R. Charon, Narrative medicine: honoring the stories of illness, Oxford University Press, Oxford; New York, 2006.
[21] R. Charon, The principles and practice of narrative medicine, Oxford University Press, New York, NY, 2017.
[22] D. Gramling, R. Gramling, Palliative care conversations: clinical and applied linguistic perspectives, Boston: De Gruyter Mouton (2019).
[23] C.K. Riessman, Narrative analysis, Sage Publications, Newbury Park, CA, 1993.
[24] A. Jaworski, N. Coupland, The discourse reader, Third edition. ed., Routledge, London ; New York, 2014.
[25] G. Lakoff, S. Narayanan, Toward Computational Model of Narrative. AAAI Fall Symposium, (2010) .
[26] W. Labov, Locating language in time and space, Academic Press, New York, 1980.
[27] D. McClure, Distributions of words across narrative time in 27,266 novels, Techne, 2017, pp. 2019. https://litlab.stanford.edu/techne/.
[28] A.J. Reagan, L. Mitchel, D. Kiley, C.M. Danforth, P.S. Dodds, The emotional arcs of stories are dominated by six basic shapes, EPJ Data Science. 5 (31) (2016).
[29] R. Gramling, E. Gajary-Coots, S. Stanek, et al., Design of, and enrollment in, the palliative care communication research initiative: a direct-observation cohort study, BMC Palliat Care. 14 (2015) 40.
[30] P.S. Dodds, K.D. Harris, I.M. Kloumann, C.A. Bliss, C.M. Danforth, Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter, PLoS One. 6 (12) (2011)e26752.
[31] P.S. Dodds, E.M. Clark, S. Desu, et al., Human language reveals a universal positivity bias, Proc Natl Acad Sci U S A. 112 (8) (2015) 2389–2394.
[32] A.J. Reagan, C.M. Danforth, B. Tivnan, J.R. Williams, P.S. Dodds, Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs, EPJ Data Science. 6 (2017).
[33] L. Mitchell, M.R. Frank, K.D. Harris, P.S. Dodds, C.M. Danforth, The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place, PLoS One. 8 (5) (2013)e64417.
[34] K. Allan, K.M. Jaszczolt, The Cambridge handbook of pragmatics, Cambridge University Press, 2012.
[35] K. Jaszczolt, Ld. Saussure, Time: language, cognition, and reality, First edition. ed, Oxford University Press, Oxford, 2013.
[36] I.D. Yalom, Existential psychotherapy, Basic Books, New York, 1980.
[37] K. Jaszczolt, Representing time : an essay on temporality as modality, Oxford University Press, Oxford; New York, 2009.
[38] R. Gramling, S. Norton, S. Ladwig, et al., Latent classes of prognosis conversations in palliative care: a mixed-methods study, J Palliat Med. 16 (6) (2013) 653–660.
[39] M. Hoerger, J.A. Greer, V.A. Jackson, et al., Defining the Elements of Early Palliative Care That Are Associated With Patient-Reported Outcomes and the Delivery of End-of-Life Care, Journal of Clinical Oncology. 36 (11) (2018) 1096–1102.
[40] D. Kavalieratos, J. Corbelli, D. Zhang, et al., Association Between Palliative Care and Patient and Caregiver Outcomes: A Systematic Review and Meta-analysis, JAMA 316 (20) (2016) 2104–2114.
[41] I.D. Yalom, Staring at the sun: overcoming the terror of death, Jossey-Bass, San Francisco, 2008.

Please cite this article in press as: L. Ross, et al., Story Arcs in Serious Illness: Natural Language Processing features of Palliative Care Conversations, Patient Educ Couns (2019), https://doi.org/10.1016/j.pec.2019.11.021