2019

Exploration of Story Arcs in Palliative Care Conversations Using Natural Language Processing

Lindsay May Ross

Exploration of Story Arcs in Palliative Care Conversations Using Natural Language Processing

Lindsay Ross

Abstract

Palliative care is an approach to improving the quality of life for patients with a serious, most likely terminal, condition. Palliative care conversations are often referred to by professionals as 'narratives', as the conversations are guided dynamically to best fit patient needs. Using transcribed text conversations of 354 palliative care consultations from 225 patients, we investigate trends in word usage over narrative time. Using crowdsourced sentiment rankings of the most common terms in the English language, we find that decreasing references to illness terms increases sentiment over narrative time. We then explore temporal references by looking at the usage of *yesterday*, *today*, and *tomorrow*, as well as variations in verb tense more generally. We find that discussion of the past decreases throughout the conversation, while discussion about the future increases. Our findings provide clinically-relevant insight into the storyline of a palliative care conversation, helping professionals to better understand these critical discussions.

Introduction

Conversations between palliative care doctors and patients are essential to ensure that patients with life-threatening illnesses receive care that is congruent with their expressed wishes. Understanding this communication is essential, yet there exists very little empirical knowledge of what goes into a conversation of this type. Not enough is being done to guarantee the success of end-of-life conversations; no practice or policy designed to improve patient life when faced with a life-threatening illness is backed by sufficient evidence.[1] The success of health-care conversations is dependent upon the quality of conversations between patients and doctors.[2] In order to understand these incredibly important conversations, we set out to characterize the temporal and emotional dynamics revealed by word choice.

---

[1] Halpern, "Toward Evidence-Based End-of-Life Care."
[2] Drew, Chatwin, and Collins, "Conversation Analysis."

The hope is that this analysis will better inform palliative care professionals about serious, end-of-life conversations.

In this analysis, we will investigate the words spoken during palliative care conversations using Natural Language Processing (NLP) techniques. NLP is key for achieving evidence-based end-of-life care.[3] Multiple studies have demonstrated the benefit of utilizing NLP technologies in healthcare, however, the majority of these studies have focused on electronic health records. Our analysis stands out from these investigations of healthcare conversations. First of all, we focus solely on palliative care, concentrating specifically on the linguistic composition of end-of-life conversations. Furthermore, the Palliative Care Communication Research Initiative (PCCRI) has been designed to provide direct observation of end-of-life conversations.[4] What sets our analysis apart from previous attempts to understand palliative care conversations, or other healthcare conversations, is the fact that we are directly analyzing the words used in the clinical dialogue, instead of just words used to describe the patient's condition.

In our analysis, we look at how the conversation changes over time, so that it can be viewed as a narrative. We look at how sentiment, word usage, and temporal reference fluctuate over time. Investigating word usage fluctuation over time has been done before – the Stanford Literary Lab explored the distribution of words when averaged across thousands of novels.[5] The analysis evaluated the shape of the narrative with the smallest unit of meaning – words. In our analysis, we will employ the narrative time concept from the Stanford Literary Lab's analysis to determine the shape of palliative care conversations.

Prior to this work, sentiment analysis has been used in hundreds of studies to quantify the emotion of large texts. For example, sentiment analysis has been used to analyze the happiness of different languages through crowdsourced scores of large, widely used texts, revealing that the human language possesses a positivity bias.[6] Additionally, sentiment analysis has previously been utilized to

---

[3] Halpern, "Toward Evidence-Based End-of-Life Care."
[4] Gramling et al., "Design of, and Enrollment in, the Palliative Care Communication Research Initiative."
[5] Stanford Literary Lab, "Distributions of Words across Narrative Time in 27,266 Novels."
[6] Dodds et al., "Human Language Reveals a Universal Positivity Bias."

determine common emotional arcs, archetypal shapes of stories.[7] Through exploration of sentiment fluctuations over narrative time, Reagan et al. found six common emotional arcs that narratives take on, demonstrating that sentiment fluctuation can indicate emotional trajectory of a narrative.

Additionally, prior to our analysis, computational linguistic studies have been done to analyze texts based upon sentence syntax. The Natural Language Toolkit (NLTK), a package used for working with human language data in Python, looks at a sentence's full structure to determine each word's part of speech.[8] In addition, a study has found that different languages vary in the way that time is encoded; some languages are inherently more forward-thinking, as there are more associations between present and future in sentence syntax.[9] Sentence structures have been analyzed to investigate verb tense as well as temporal reference relationships. In our analysis, we utilize NLTK's method of tagging parts of speech, and further investigate the sentence syntax to determine temporal reference.

In our analysis, we first explore the sentiment of the words used over time in palliative care conversations, and investigate what determines the change in conversation sentiment as narrative time passes. Next, we explore how words relating to patient illness – including symptoms, treatments, and prognosis – fluctuate over time. We then examine how discussion of the past, present, and future vary during the conversation. Lastly, we look into the dynamics of modal verbs, indicating possibility and probability.

Methods

Overview

This is a cross-sectional analysis of 354 transcriptions of audio recorded inpatient palliative care consultations. We analyzed word usage over narrative time using NLP techniques.

[7] Reagan et al., "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes."
[8] Loper and Bird, "NLTK."
[9] Chen, "The Effect of Language on Economic Behavior."

Data

As described more fully elsewhere[10], the PCCRI is a multisite observational cohort study, conducted between January 2014 and May 2016, where 240 patients with advanced cancer were enrolled at the time of referral for palliative care conversations. From these patients, four withdrew, three passed away, and two were discharged before the study was complete, therefore data from 231 patients is utilized. Each patient participated in one, two, or three palliative care conversations, each of which was recorded, transcribed verbatim, then prepared for computational processing.

Conversations in the dataset had varying lengths, with a median word count of 3352.5. For reference, to the right is a histogram displaying the different conversation word counts (Figure 1).

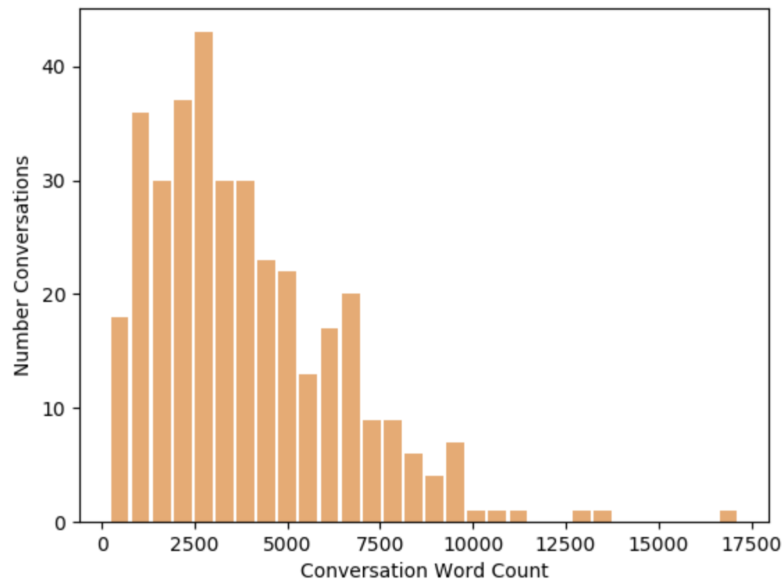


Figure 1 – This figure displays a histogram of the conversation word counts for each conversation in our dataset.

Measures

In our analysis, we assess the sentiment of a palliative care conversation using a roughly 10,000 word sentiment dictionary. This list was created by crowdsourcing participants in an online marketplace created by Amazon called Mechanical Turk. The language assessment by Mechanical Turk (labMT), described in detail elsewhere[11], was developed by first combining the most frequently used 5,000 words found in tweets, New York Times articles, Google Books, and music lyrics. 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

---

[10] Gramling et al., "Design of, and Enrollment in, the Palliative Care Communication Research Initiative."
[11] Dodds et al., "Temporal Patterns of Happiness and Information in a Global Social Network."

example, the words “worse”, “of”, and “happy” received average scores of 2.77, 4.94, and 8.30 respectively.

In our exploratory analysis of trends in word usage, we noticed changing frequencies of clinical terms describing patient illness, specifically terms discussing symptoms, treatments, and prognosis related to illness. 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 these groupings, we only considered words used more than 100 times in the conversation data. There are 17,041 unique words in our dataset, and thresholding above 100 left 947 unique words, enabling a focus on the most reliable trends. We manually categorized words as belonging to one of the following categories if exactly one of the definitions were met:

Symptom Terms – terms used exclusively to discuss what patient is experiencing due to illness

Treatment Terms – terms used exclusively to discuss what can be done to improve patient symptoms and illness

Prognosis Terms – terms used exclusively to discuss patient future in relation to their illness

See Appendix A for a complete list of symptom, treatment, and prognosis terms used. To further investigate how these terms fluctuated together, we lumped all symptom, treatment, and prognosis terms together, and refer to them as illness terms.

To investigate how trends in temporal reference fluctuated over time, we developed a method for determining whether an individual verb usage was referencing the past, present, or future. There are multiple different verb forms and verb tenses that a verb can take on in the English language. These can be identified by looking into how the verb is conjugated, and the words that precede that conjugated verb. To do this, we processed data from NLTK. This package, when given a text, tags each word with a part of speech. We use NLTK to identify where a verb is used, and how it is conjugated. Looking at this conjugation, and the words that appear before that conjugated verb, we determine whether the verb is temporally referencing the past, present or future. See Appendix B for further information on how NLTK tags verbs, and for information on which verb form/tense is grouped into each temporal reference category.

Lastly, throughout the analysis, when we refer to palliative care conversations, we are referring to the conversations not solely between patients and doctors, but between the patient-side (meaning the patient and all present family members) and the doctor-side (meaning all clinical staff present).

Analytic Approach

When analyzing trends in word usage over time, we employ the concept of 'narrative time'[12], where each conversation's timeline is normalized to a dimensionless axis representing the percentage of the conversation that has occurred up to that moment. Since the number of words used in conversations in our data varies from 196 to 17,166, the number of words per percentage also varies. 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, we split the conversation into deciles of narrative time. While the choice of 10 bins is somewhat arbitrary, a sensitivity study on the number of bins demonstrated that the trends we saw were robust to changes in units of dimensionless time. Appendix C presents this sensitivity study.

We analyze word usage by exploring the relative frequency of the words used with respect to the background corpus of palliative care conversations. In other words, for each decile of the conversation, we calculate the number of times the word(s) of interest appear in that specific decile in all conversations, divided by the number of times the word(s) of interest appears in total over all conversations.

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.

[12] Stanford Literary Lab, "Distributions of Words across Narrative Time in 27,266 Novels."

Results

Sentiment Analysis

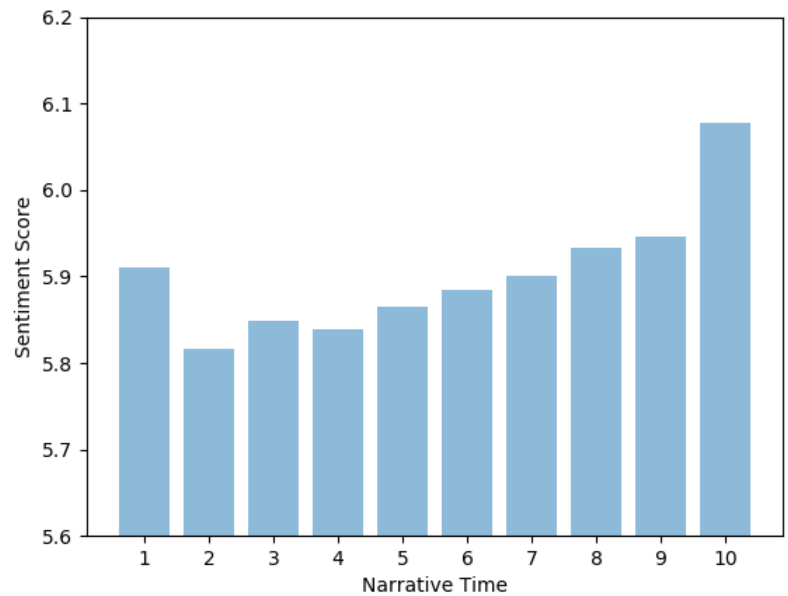


Figure 2 – This figure shows the fluctuation in average sentiment over the deciles of narrative time. Average sentiment calculated by labMT word scores.

By calculating the average sentiment of palliative care conversations in each decile, we find that palliative care conversations typically increase in sentiment as narrative time passes (Figure 2). Moving through each decile of narrative time, we note that the conversation sentiment score starts in the first decile at 5.91 and drops to 5.82 in the second decile, then increases over the following deciles, reaching 6.08 in the final decile. To put this change in sentiment into perspective, we reference the average sentiment of 10% of all tweets on specific days using the Hedonometer[13]. Similar to the first decile, the 2017 terrorist attack on Barcelona had a sentiment score of 5.92. Like the second decile, the 2016 mass shooting of Pulse nightclub in Orlando had a sentiment score of 5.84. When we reach the last decile, this is similar to the sentiment score of U.S. holidays – Easter 2017 had an average sentiment of 6.08, and Mother's Day 2018 had a score of 6.09. The change in sentiment over time in palliative care conversations thus moves from strongly negative to strongly positive.

[13] University of Vermont Computational Story Lab, "Average Happiness for Twitter." http://hedonometer.org

To look into why this sentiment change occurs, we investigate the words that contribute the most to the difference in sentiment scores between deciles using figures called ‘word shift graphs’. First, we investigate the drop in sentiment that occurs from the first to the second decile (Figure 3). The word shift displays why the first decile is happier, on the right side of the shift, through showing the “happy” words used more and the “sad” words used less. The first decile of conversations uses the relatively positive words like “okay”, “good”, and “nice” more often and uses relatively sadder negation words such as “not”, “don’t”, “doesn’t”, “didn’t”, and “no” less often. Also, the first decile contains the words “hello” and “hi” more often than the second decile. Additionally, the first decile is more positive than the second decile due to the lack of negative words discussing the illness such as “cancer” and “pain”, as well as “blood”, “surgery”, and “death” (not shown). In Figure 3, only the first 23 words are displayed for brevity, but the shift in full provides us access to the top 200 words that drive this sentiment difference.

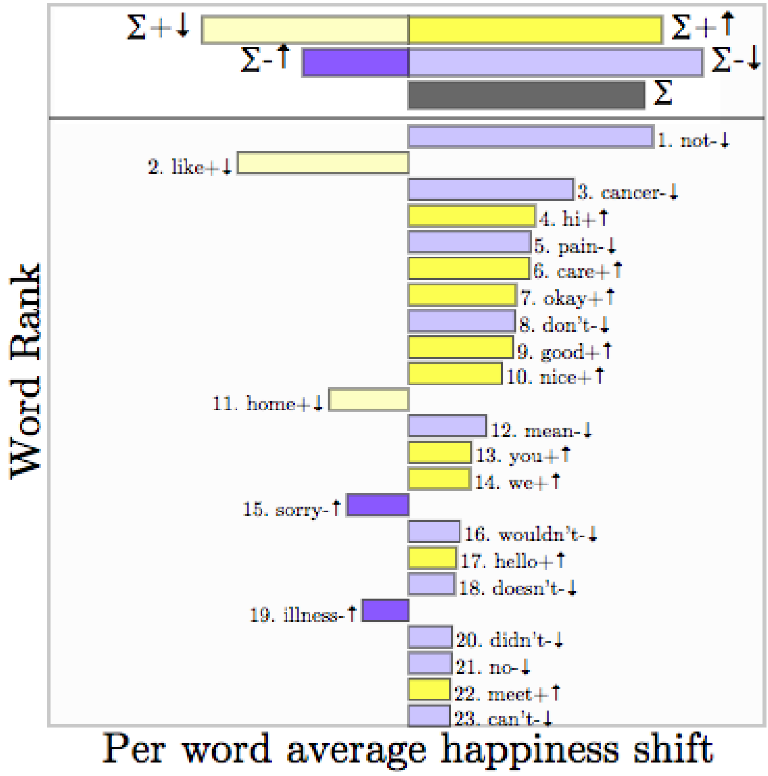


Figure 3 – This word shift shows why the first decile of a palliative care conversation is happier than the second decile. The figure displays the words driving the difference between the first and the second decile, in order of contribution to the decile’s sentiment score. The right side represents, when compared to the second decile, what contributes to the increase in happiness of the first decile, and the left represents what contributes to the decrease in happiness of the first decile. The + and – symbol indicate whether a word used was relatively happy or sad, while the up and down arrows indicate whether a word was used more or less frequently.

As we move through narrative time, from the second to the final decile, a series of word shifts indicate that the sentiment score difference between consecutive deciles is primarily driven by changes in the frequency of terms specifically discussing illness. Screenshots of the word shifts of consecutive deciles can be found in Appendix D. For example, from the second decile to the third decile, we see a decrease in the terms "pain", "cancer", "severe", and "failure", contributing to an increase in sentiment. When we move from the third to the fourth decile, we see a decrease in terms "cancer", "radiation", "surgery", and "sick", helping to drive an increase in sentiment. Moving from the fourth to the fifth decile, the terms "pain", "hospital", and "surgery" decrease, again helping to cause a positive sentiment change. Patterns like this continue as the conversation continues, providing insight into the type of words driving the increase in sentiment from the second decile to the end of the conversation. To summarize the words responsible for the steady sentiment increase, Figure 4 displays the word shift showing why the ninth decile is happier than the second.

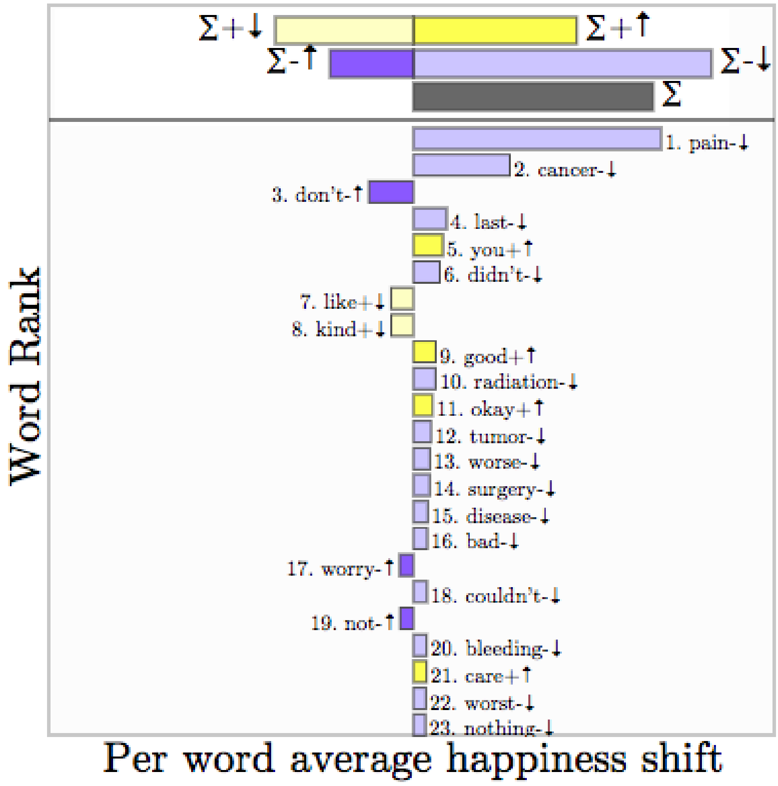


Figure 4 – This word shift shows why the ninth decile of a palliative care conversation is happier than the second decile. The figure displays the words driving the difference between the ninth and the second decile, in order of contribution to the decile's sentiment score. The right side represents, when compared to the second decile, what contributes to the increase in happiness of the ninth decile, and the left represents what contributes to the decrease in happiness of the ninth decile. The + and – symbol indicate whether a word used was relatively happy or sad, while the up and down arrows indicate whether a word was used more or less frequently.

When investigating why there is a larger jump in positive sentiment at the end of these conversations, we see that the difference is again driven by terms discussing the illness. Moving from the ninth decile to the tenth decile, we see a decrease in illness related terms "pain", "cancer", "hospital", and "hurt", helping to drive an increase in sentiment. We also notice an increase in words expressing thankfulness when we reach the last decile, including words such as "thank", "nice", "good", and "pleasure". Lastly, we notice a decrease in terms of negation "not" and "don't".

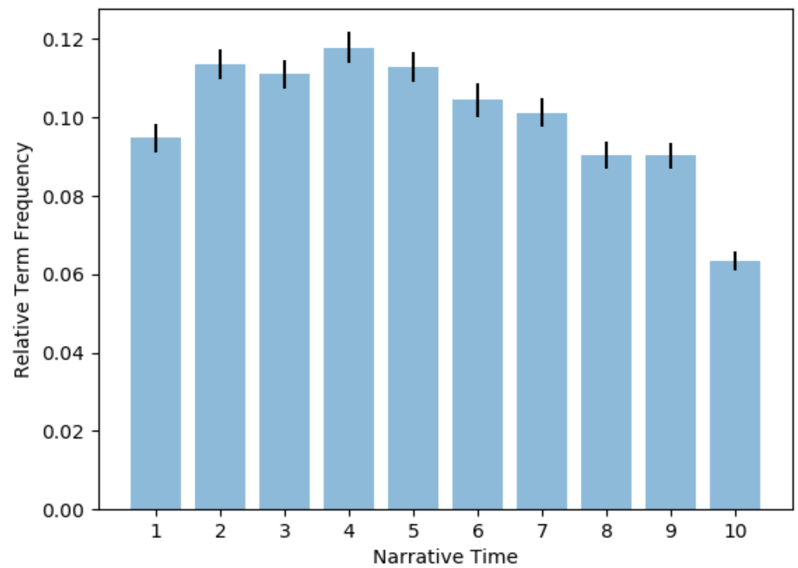


Figure 5 – Relative frequency of illness terms in each decile of the conversation in narrative time, with standard error bars displayed. In total, illness terms appear 35,624 times in the palliative care conversation data.

Discussion of Illness

The relative frequency of all illness terms (consisting of words related to symptoms, treatments, and prognosis) increases at the start of conversations but then decreases from the fourth decile to the final decile (Figure 5).

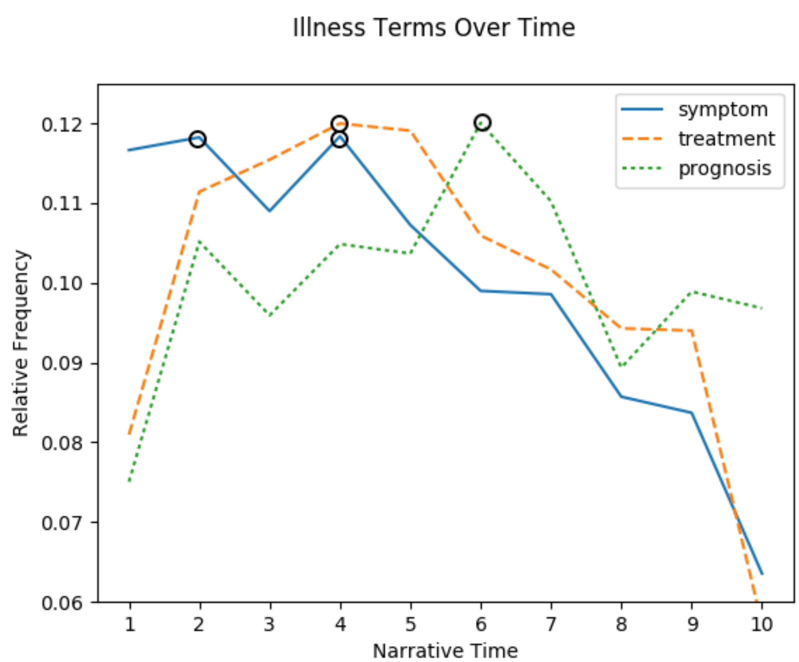


Figure 6 – Fluctuation in relative frequency of symptom, treatment, and prognosis terms over each decile of narrative time. In our palliative care conversation data, symptom terms appear 14,396 times, treatment terms appear 17,871 times, and prognosis terms appear 3,357 times. Peaks referenced in the document are circled.

Disaggregating illness terms used in the conversation into symptom, treatment, and prognosis terms, we see that each peaks at successively later times in the conversation (Figure 6). Specifically, the two highest relative frequencies of symptom terms are at the second decile (0.1182) and the fourth decile (0.1184). Following this, treatment terms peak in the fourth decile (0.1200), and prognosis terms peak in the sixth decile (0.1200).

Temporal Reference Analysis

The discussion of the past, present, and future also varies throughout narrative time. One indication of this change is the relative frequency of the terms “yesterday”, “today” and “tomorrow” over time. We observe that the usage of “yesterday” generally decreases, the usage of “today” dips in the middle of the conversation, and the usage of “tomorrow” increases as the conversation progresses, particularly in the final decile (Figure 7).

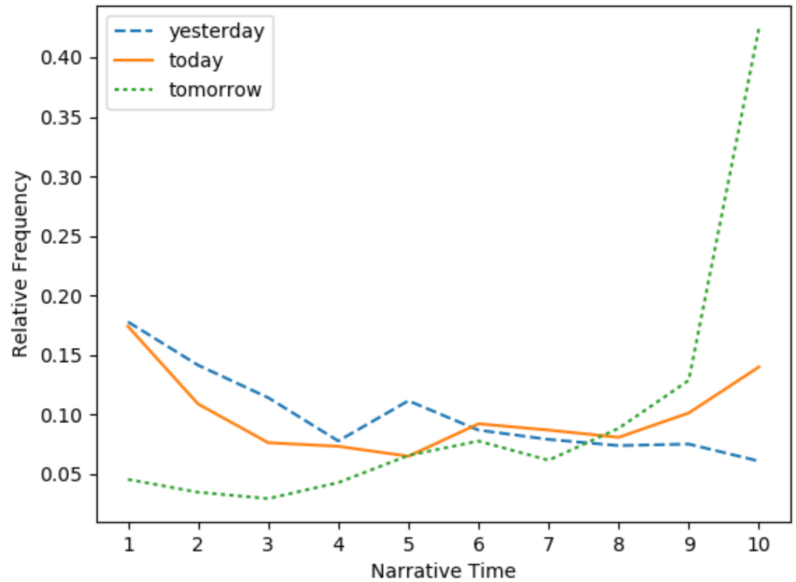


Figure 7 – This figure displays the relative frequency of temporal nouns “yesterday”, “today”, and “tomorrow” fluctuating over the deciles of narrative time. In all palliative care conversations, “yesterday” occurs 770 times, “today” occurs 1,321 times, and “tomorrow” occurs 745 times.

We also investigate the variation between past, present, and future through the temporal reference of the verbs used throughout the conversation, and observe a strong difference in the shapes of the three categories of temporal reference (Figure 8). We observe the relative frequency

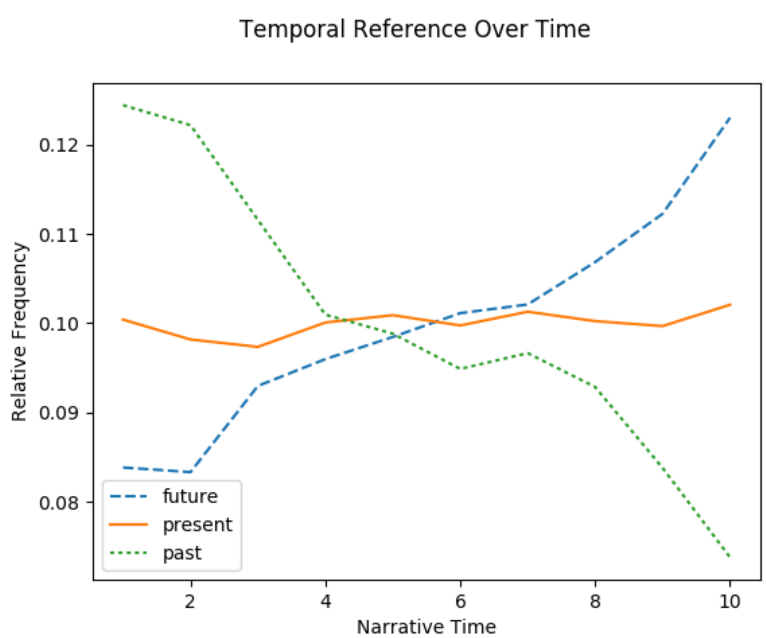


Figure 8 – This figure displays the relative frequency of temporal reference in each conversation fluctuating over the deciles of narrative time. In all palliative care conversations, according to our method, temporal reference of the past occurs 63,913 times, temporal reference of the present occurs 181,662 times, and temporal reference of the future occurs 54,473 times.

of discussion of the past decreases sharply over narrative time, discussion of the present is relatively constant with a slight upward drift over narrative time, while discussion of the future increases sharply over narrative time. These findings are consistent with, but even more dramatic than, the patterns observed for usage of "yesterday", "today", and "tomorrow".

Finally, we examine trends in the use of modal verbs. Modal verbs are auxiliary verbs utilized to express necessity or possibility; they are used to show whether or not we believe something is certain, probable or possible. The modal verbs of the English language are "can", "could", "may", "might", "shall", "should", "will", "would", and "must". We see that the relative frequency of modal verbs increases over narrative time (Figure 9).

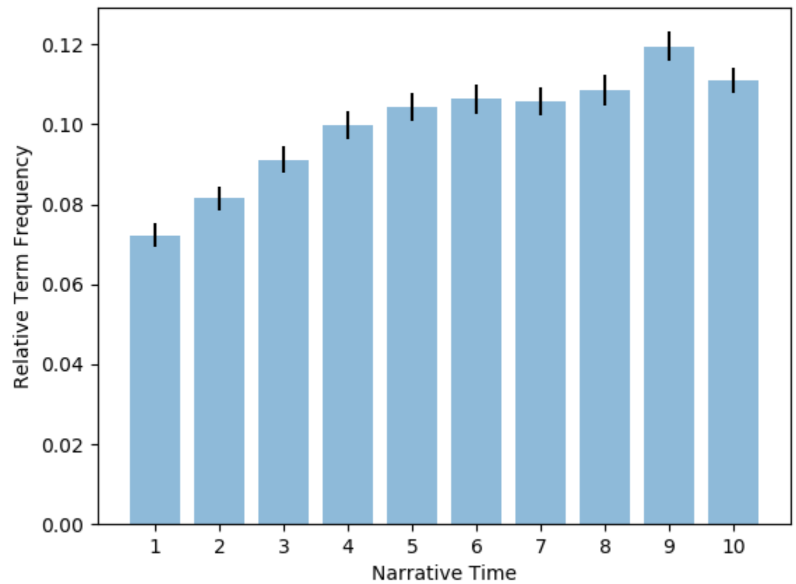


Figure 9 – Relative frequency of modal verbs in each conversation decile of narrative time. In our conversation data, modal verbs appear 21,282 times.

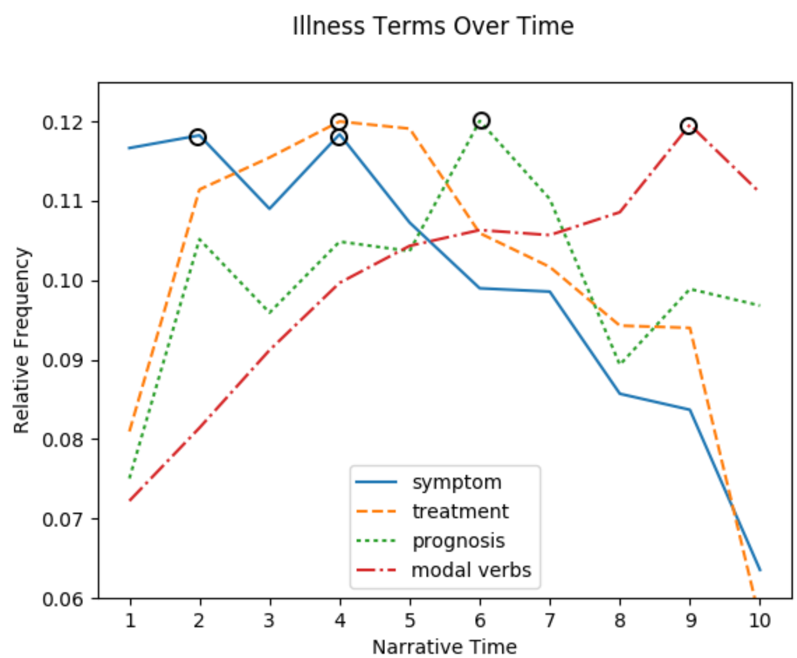


Figure 10 – Fluctuation in relative frequency of symptom terms, treatment terms, prognosis terms, and modal verbs over each decile of narrative time. Peaks referenced in the document are circled.

When we investigate how the discussion of modal verbs fits in with the disaggregated illness terms, we see that the peak of modal verbs falls after prognosis terns, peaking in the ninth decile (Figure 10).

Discussion

Our results show that the fluctuation of sentiment over time in palliative care conversations appears to be primarily driven by the usage of illness terms. The sentiment of the first decile starts off higher than the second decile. We can speculate that this is due to the fact the doctor and patient have not yet begun discussing the illness. Then, the sentiment of the conversation dips in the second decile as the conversation is directed towards the illness. The sentiment then increases from the second decile of the conversation to the end, driven by the continual decrease of the discussion of illness. When we look beyond sentiment, and further analyze illness discussion by exploring how illness terms fluctuate over time, we see results consistent with our expectations – discussion of illness increases slightly at the beginning of narrative time and then decreases, with a natural progression from discussion of symptoms, to treatments, to prognosis.

From analysis of the temporal nouns “yesterday”, “today”, and “tomorrow”, and of the temporal reference of the verbs used, we see that as time progresses, patient-clinician discussion of the past decreases over time, while discussion of the future increases. Modal verbs, indicating belief in whether something is certain, probable, or possible, continually increase over time. This suggests that the increase in discussion of the future may be caused, in part, by the discussion of probability and possibility of future events. The observation that discussion of prognosis (which anticipates the patient’s future in relation to their illness) peaks relatively late in these conversations provides another possible indication as to why temporal reference of the future increases over time. Additionally, when we see how modal verbs peak after peaks of symptom, treatment, and prognosis terms, we can suppose that discussion of future probabilities and possibilities takes place after the discussion of the illness.

The goal of this analysis was to explore the trends in word usage found in palliative care conversations, which we achieved by looking into how word groupings, the sentiment of the words used, and the temporal reference of the words used tended to fluctuate over time. Moving forward, it would be interesting to investigate trends on an individual conversation level. Based on survey data recorded before and after patient conversation(s), we have indications on the degree to which palliative care conversations made individual patients feel “heard and understood” by clinicians. In the future, we can explore word usage trends in individual conversations and investigate how those trends correlate with how heard and understood the patient ultimately felt due to the conversation. Additionally, we have data, from the PCCRI survey on patients' self-reported level of optimism, and their self-reported quality-of-life before and after the palliative care conversation(s). Another interesting future goal for this research would be to see how sentiment fluctuates in more optimistic patients in comparison to more pessimistic patients, and whether there is an association between increases in positive sentiment and increases in quality-of-life.

Conclusion

Patient-doctor communication in palliative care conversations is essential to making patients feel heard and understood at the end of their life, thus helping to ensure that their care is concordant with their wishes. This analysis provides empirical evidence of what goes into these extremely important conversations. According to our dataset, we see multiple trends in palliative care conversations. The positive sentiment of a palliative care conversation generally increases as narrative time passes, largely due to the decrease in the usage of symptom, treatment, and prognosis terms that characterize the patient illness. Additionally, we see how discussion of the past decreases, while discussion of the future increases. We can speculate that the increase in discussion of the future is correlated with the increase in modal verbs, indicating probability in the future, and the increase of prognosis terms, discussing the future in relation to the patient's illness. Overall, this analysis may provide palliative care clinicians with a better understanding of dynamics in the narrative of palliative care conversations.

Acknowledgements

This work was funded, in part, by a Research Experience for Undergraduates Award from the University of Vermont College of Engineering and Mathematical Sciences and a Scholarship funded by the National Science Foundation (S-STEM award, DUE-1356426, PI: Margaret J. Eppstein), and would not have been possible without the data provided from the PCCRI, which was funded by a Research Scholar Grant from the American Cancer Society (RSG PCSM124655; PI: Robert Gramling). In addition, this work would not have been possible without my thesis advisors, Margaret J. Eppstein and Chris Danforth, my additional thesis committee members, Donna Rizzo and Robert Gramling, and the other members of the Vermont Conversation Lab: Laurence Clarfeld, Brigitte Durieux, Laura Hirsch, and Cailin Gramling. Lastly, the analysis would not have been possible without those who helped prepare the transcripts for computational processing – Aidan Ryan, James Cazayoux, and Bridger Banco.

References

Chen, M. Keith. "The Effect of Language on Economic Behavior: Evidence from Savings Rates, Health Behaviors, and Retirement Assets." *American Economic Review* 103, no. 2 (April 2013): 690–731. https://doi.org/10.1257/aer.103.2.690.

Dodds, Peter Sheridan, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, et al. "Human Language Reveals a Universal Positivity Bias." *ArXiv:1406.3855 [Physics]*, June 15, 2014. http://arxiv.org/abs/1406.3855.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter." Edited by Johan Bollen. *PLoS ONE* 6, no. 12 (December 7, 2011): e26752. https://doi.org/10.1371/journal.pone.0026752.

Drew, Paul, John Chatwin, and Sarah Collins. "Conversation Analysis: A Method for Research into Interactions between Patients and Health-Care Professionals." *Health Expectations* 4, no. 1 (March 2001): 58–70. https://doi.org/10.1046/j.1369-6513.2001.00125.x.

Gramling, Robert, Elizabeth Gajary-Coots, Susan Stanek, Nathalie Dougoud, Heather Pyke, Marie Thomas, Jenica Cimino, et al. "Design of, and Enrollment in, the Palliative Care Communication Research Initiative: A Direct-Observation Cohort Study." *BMC Palliative Care* 14, no. 1 (December 2015). https://doi.org/10.1186/s12904-015-0037-8.

Halpern, Scott D. "Toward Evidence-Based End-of-Life Care." *New England Journal of Medicine* 373, no. 21 (November 19, 2015): 2001–3. https://doi.org/10.1056/NEJMp1509664.

Loper, Edward, and Steven Bird. "NLTK: The Natural Language Toolkit." *ArXiv:Cs/0205028*, May 17, 2002. http://arxiv.org/abs/cs/0205028.

Reagan, Andrew J, Lewis Mitchell, Dilan Kiley, Christopher M Danforth, and Peter Sheridan Dodds. "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes." *EPJ Data Science* 5, no. 1 (December 2016): 31. https://doi.org/10.1140/epjds/s13688-016-0093-1.

Stanford Literary Lab. "Distributions of Words across Narrative Time in 27,266 Novels," n.d.

University of Vermont Computational Story Lab. "Average Happiness for Twitter," n.d. http://hedonometer.org.

Appendix

Appendix A – Symptom, Treatment, and Prognosis Terms

Symptom Terms – terms used exclusively to discuss what patient is experiencing due to illness

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

Treatment Terms – terms used exclusively to discuss what can be done to improve patient symptoms and illness

antibiotics, ativan, button, chemo, chemotherapy, cpr, dialysis, dilaudid, dose, doses, drug, drugs, feeding, fentanyl, fluids, hospice, icu, iv, line, liquid, machine, management, medical, medication, medications, medicine, medicines, meds, methadone, mg, milligrams, mm, morphine, nutrition, oral, oxycodone, oxygen, patch, pill, pills, procedure, radiation, resuscitation, surgery, therapy, treat, treatment, treatments, trial, tube, tylenol, ventilator

Prognosis Terms – terms used exclusively to discuss patient future in relation to their illness

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

Appendix B – Temporal Reference Categorization

Here, we provide a further explanation on how we determine temporal reference. We can figure out the conjugation of a verb with NLTK's part of speech tagger, tagging verbs as the following types.

TAG	PART OF SPEECH	EXAMPLE
VB	base form	"take"
VBD	past tense	"took"
VBG	gerund/present participle	"taking"
VBN	past participle	"taken"
VBP	singular present, non-third-person	"take"
VBZ	third person singular present	"takes"

Based on these tags, we know the verb's conjugation, and by looking at the words prefacing that verb, we can identify the verb tense or verb form that this tagged verb is taking on. For example, every time a VBG verb is prefaced by "was" or "were", we count this as past continuous, therefore we say that the past was temporally referenced. We place these verb tenses/forms in the following categories:

PAST

TENSE/FORM	EXAMPLE
PAST CONTINUOUS	"We were *taking*…"
PAST PERFECT CONTINUOUS	"I had been *taking*…"
PRESENT PERFECT	"She has *taken*…"
PAST PERFECT	"I had *taken*…"
PERFECT PARTICIPLE	"Having *taken* this exam before…"
PAST SIMPLE/PRETERITE	"He *took*…"

PRESENT

TENSE/FORM	EXAMPLE
PRESENT CONTINUOUS	"I am *taking*…"
PRESENT PERFECT CONTINUOUS	"I have been *taking*…"
PRESENT PARTICIPLE	"Stop *taking*…"
PRESENT SIMPLE*	"She *takes*…"

FUTURE

TENSE/FORM	EXAMPLE
FUTURE SIMPLE	"I will *take*…"
INFINITIVE	"I want to *take*…"
IMPERATIVE*	"Let's *take*…"
FUTURE CONTINUOUS	"I will be *taking*…"
FUTURE PERFECT CONTINUOUS	"I will have been *taking*…"

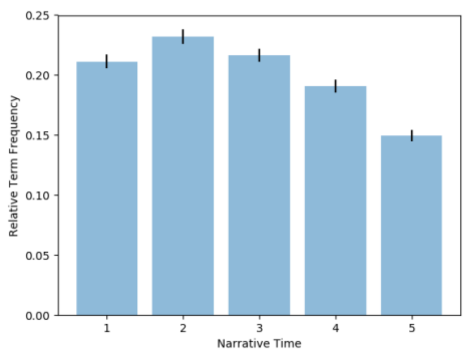
FUTURE PERFECT	“I will have *taken*…”

*NOTE – we are unable to capture all instances of the present simple and the imperative, due to the lack of a systematic way to look at the words used before a verb type and determine whether the tense/form is either present simple or imperative. For example, when solely looking at the words surrounding “*take*” in the sentences “Everyday, you *take* a seat” (present simple) and “I recommend you *take* a seat” (imperative), there is no systematic way to differentiate between the present simple and imperative based on words prefacing the verb. Therefore, to capture many, but not all, occurrences of the present simple, we categorize all VBZ and VBP verbs as present simple, therefore in the present temporal reference category. We categorize any VB verb that is prefaced by “let us” or “let’s”, as imperative, therefore in the future temporal reference category.

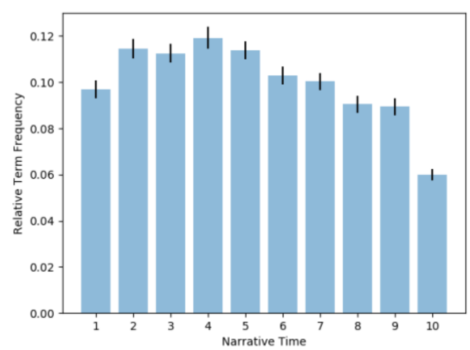
Appendix C – Bin Sensitivity Study Graphs

We performed a sensitivity study on the number of bins used in our analysis, graphing word group trends with 5, 10, 15, 20, 25, and 30 bins. Our study demonstrated that the trends we saw were robust to changes in units of dimensionless time. Here, we see this study on illness terms and modal verbs as an example.

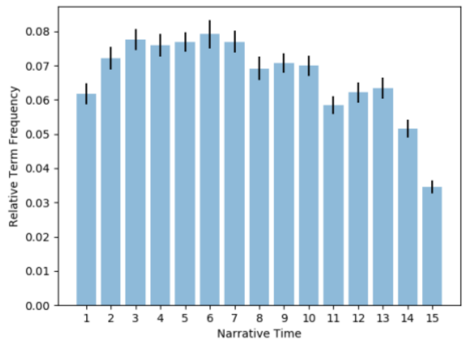
Illness Terms



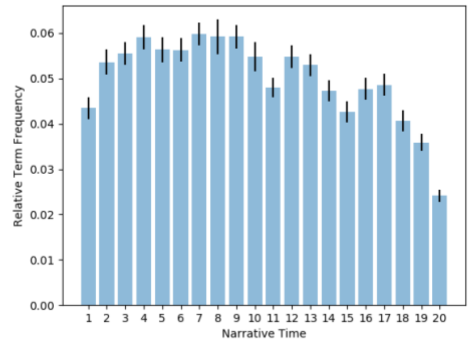
The relative frequency of illness terms over 5 bins of narrative time, with standard error bars shown.



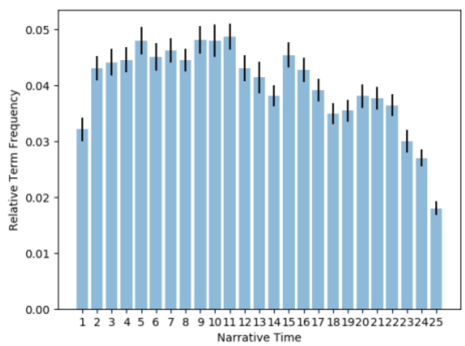
The relative frequency of illness terms over 10 bins of narrative time, with standard error bars shown.



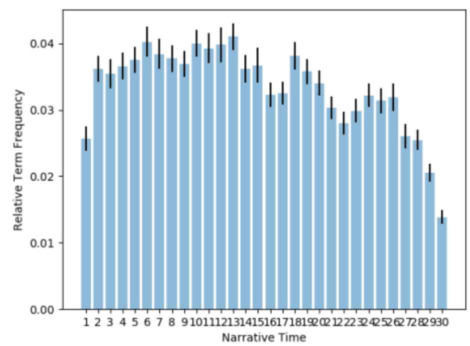
The relative frequency of illness terms over 15 bins of narrative time, with standard error bars shown.



The relative frequency of illness terms over 20 bins of narrative time, with standard error bars shown.

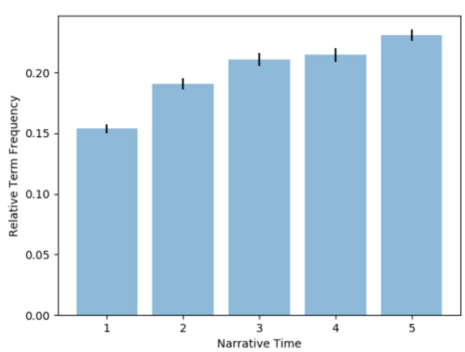


The relative frequency of illness terms over 25 bins of narrative time, with standard error bars shown.

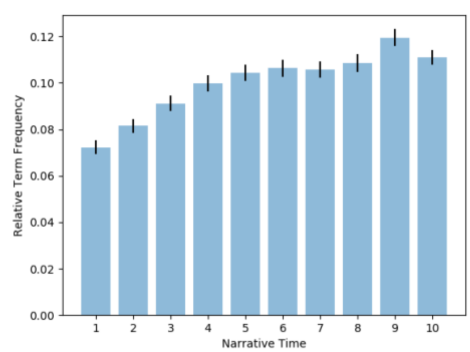


The relative frequency of illness terms over 30 bins of narrative time, with standard error bars shown.

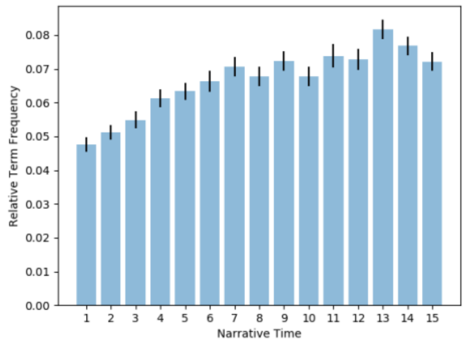
Modal Verbs



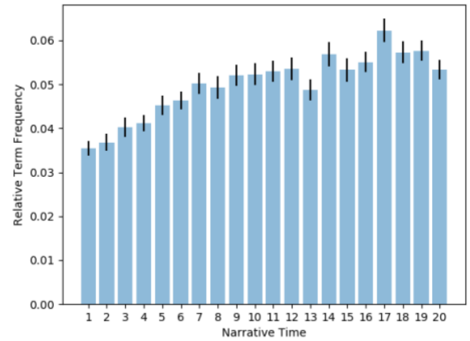
The relative frequency of modal verbs over 5 bins of narrative time, with standard error bars shown.



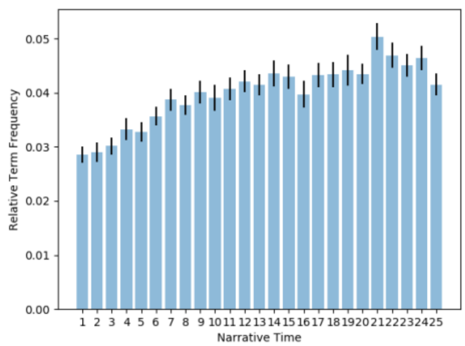
The relative frequency of modal verbs over 10 bins of narrative time, with standard error bars shown.



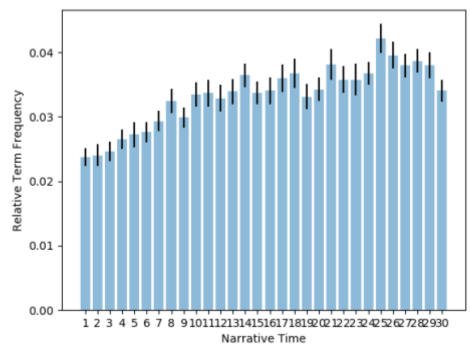
The relative frequency of modal verbs over 15 bins of narrative time, with standard error bars shown.



The relative frequency of modal verbs over 20 bins of narrative time, with standard error bars shown.



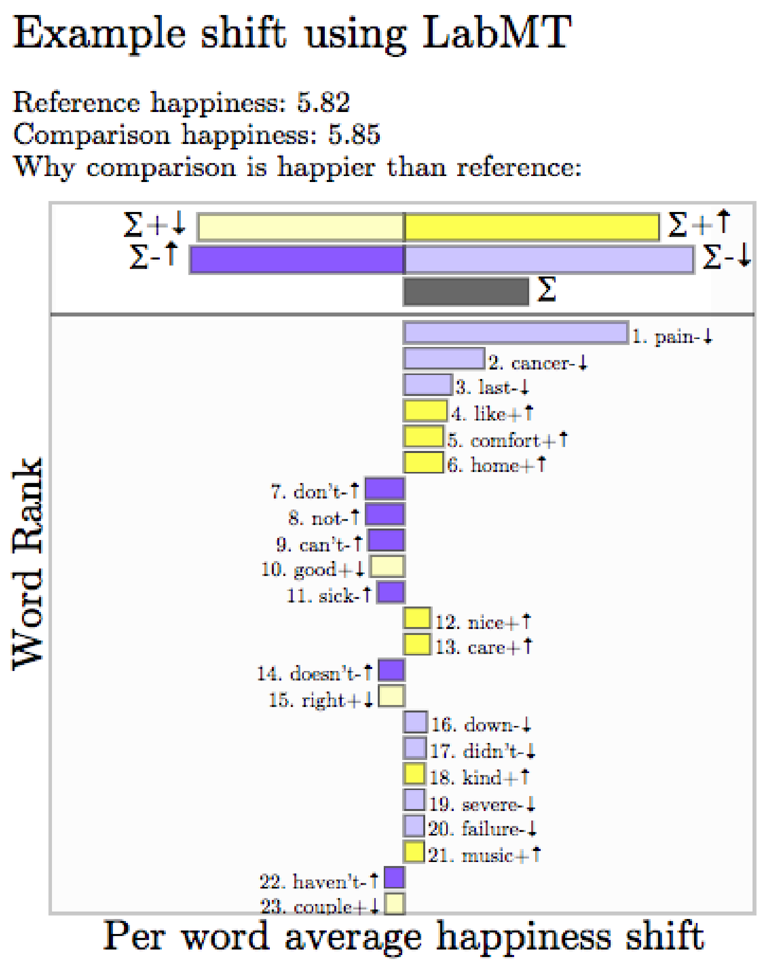
The relative frequency of modal verbs over 25 bins of narrative time, with standard error bars shown.



The relative frequency of modal verbs over 30 bins of narrative time, with standard error bars shown.

Appendix D – Word shifts of Consecutive Deciles

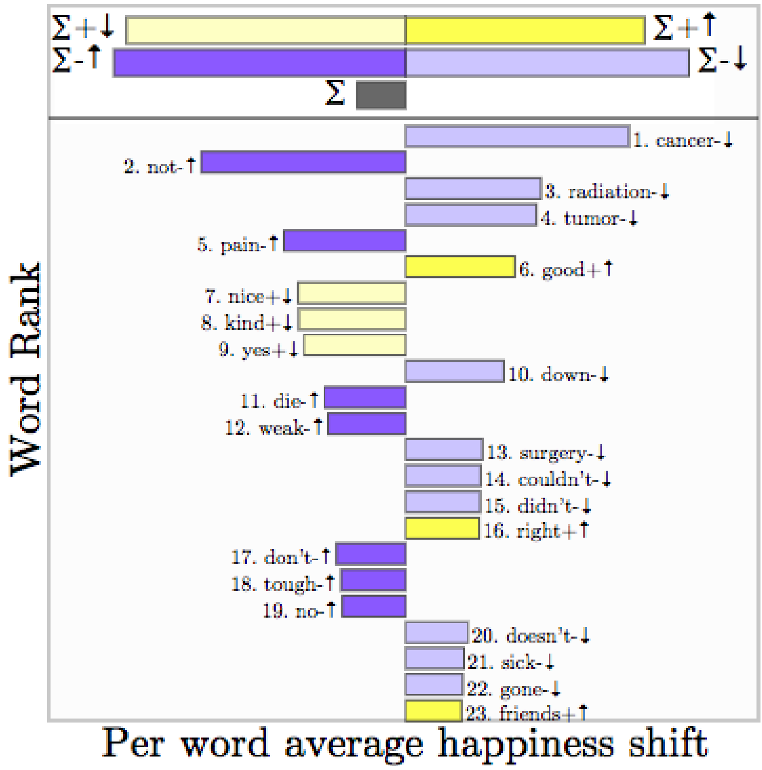
Here, we can see the word shifts of consecutive deciles, demonstrating the terms driving the difference in the change in sentiment over narrative time. When word shifts are generated, the user can scroll down and view the top 200 words driving the sentiment difference between deciles. For brevity, we will show screenshots here of the top 23 words affecting the sentiment difference.



This word shift shows why the second decile of a palliative care conversation is happier than the third decile.

Example shift using LabMT

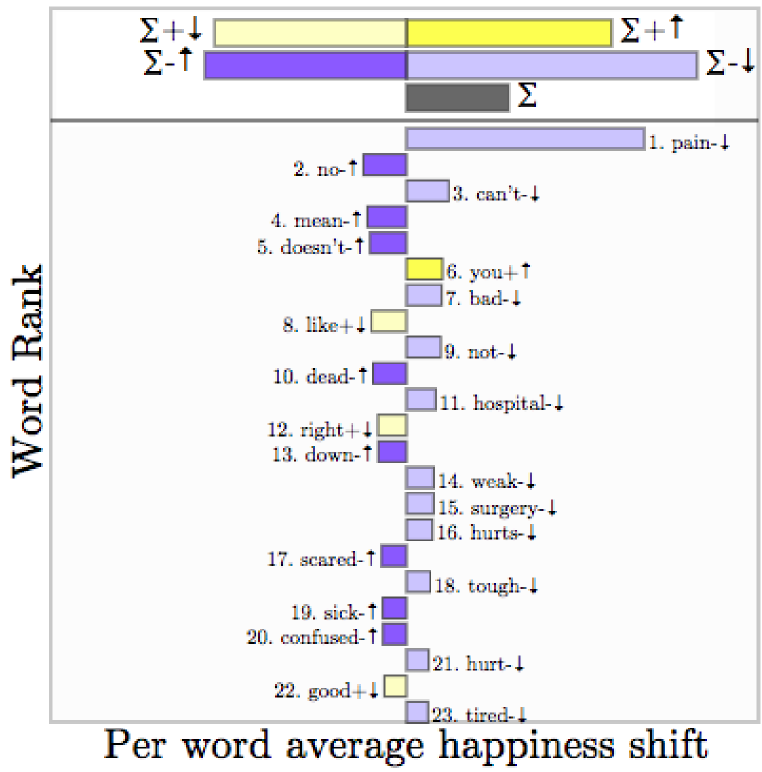
Reference happiness: 5.85
Comparison happiness: 5.84
Why comparison is less happy than reference:



This word shift shows why the fourth decile of a palliative care conversation is less happy than the third decile.

Example shift using LabMT

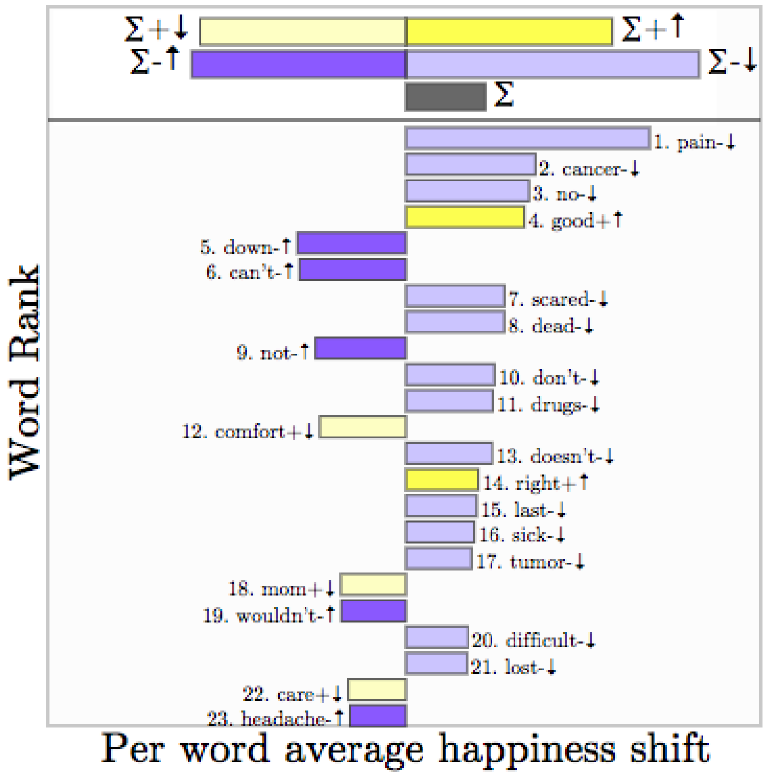
Reference happiness: 5.84
Comparison happiness: 5.87
Why comparison is happier than reference:



This word shift shows why the fifth decile of a palliative care conversation is happier than the fourth decile.

Example shift using LabMT

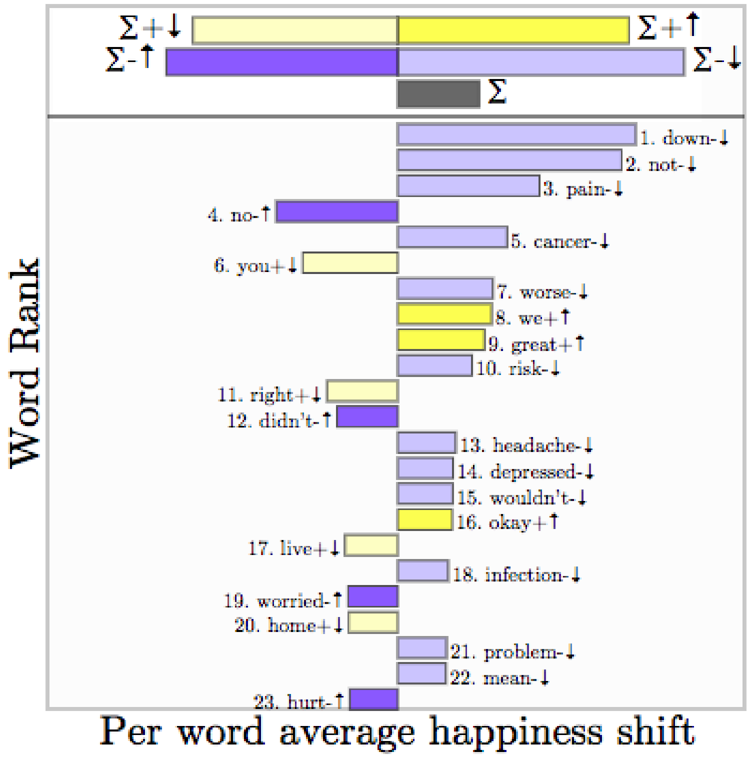
Reference happiness: 5.87
Comparison happiness: 5.88
Why comparison is happier than reference:



This word shift shows why the sixth decile of a palliative care conversation is happier than the fifth decile.

Example shift using LabMT

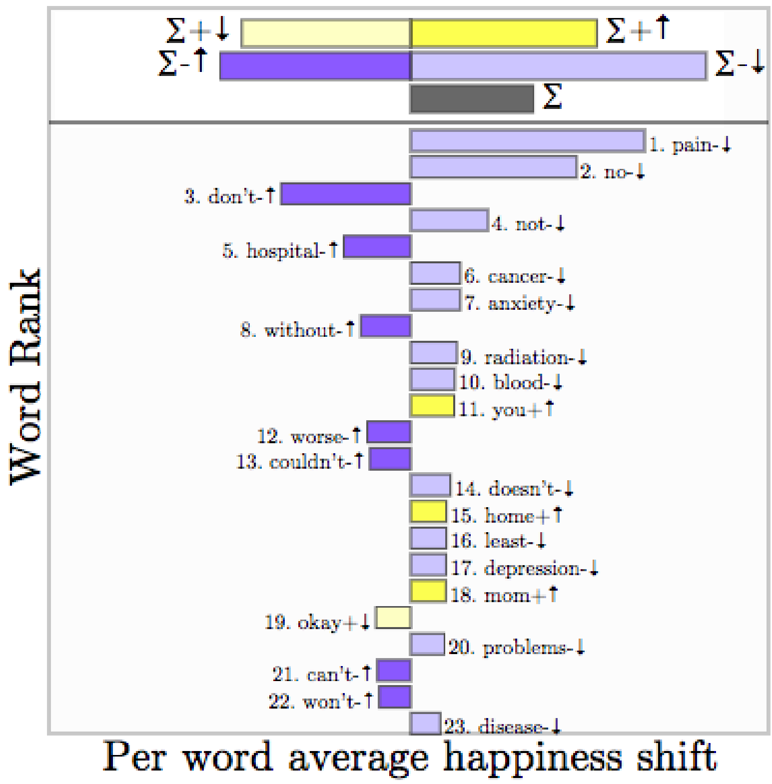
Reference happiness: 5.88
Comparison happiness: 5.90
Why comparison is happier than reference:



This word shift shows why the seventh decile of a palliative care conversation is happier than the sixth decile.

Example shift using LabMT

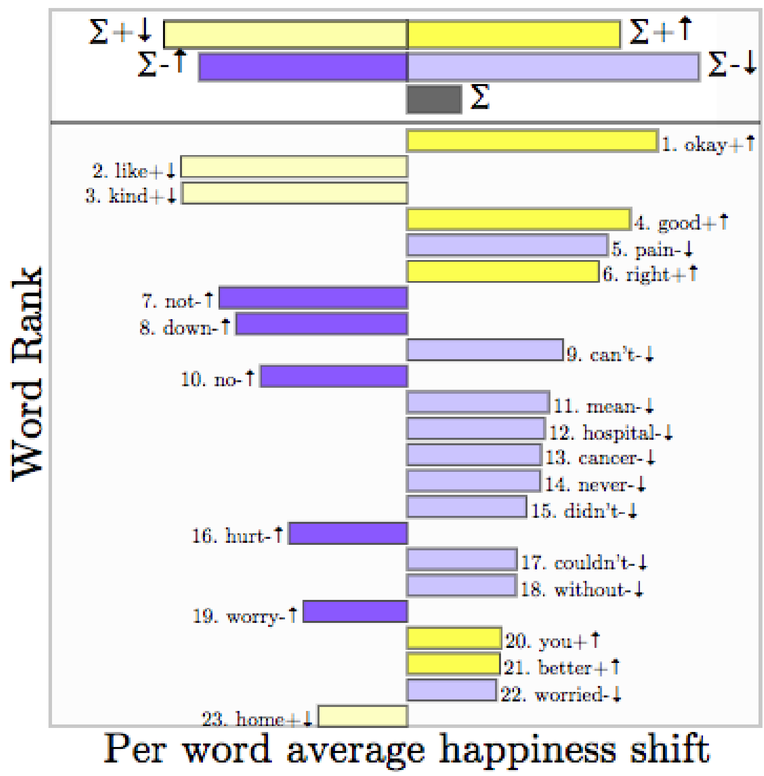
Reference happiness: 5.90
Comparison happiness: 5.93
Why comparison is happier than reference:



This word shift shows why the eighth decile of a palliative care conversation is happier than the seventh decile.

Example shift using LabMT

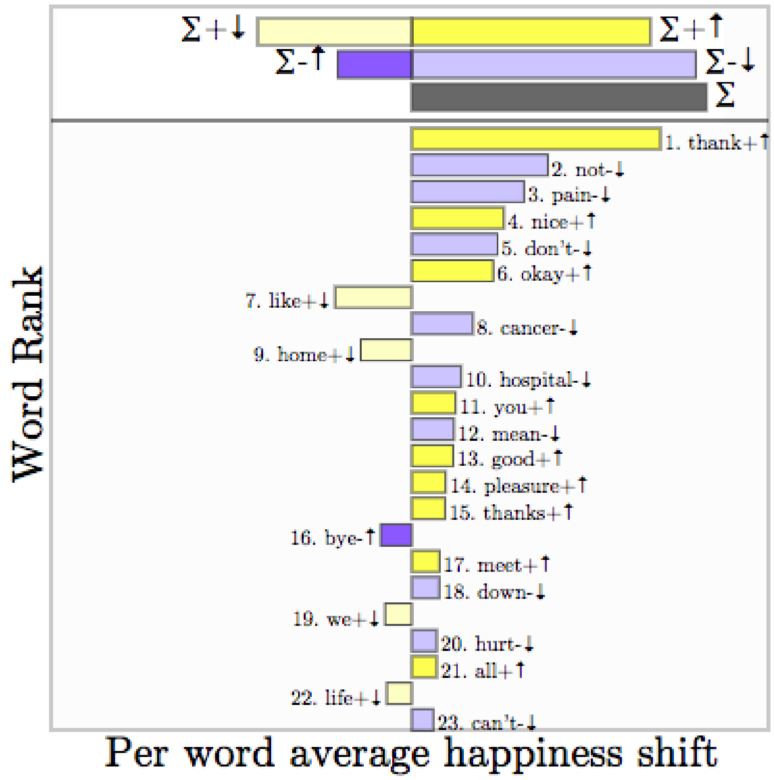
Reference happiness: 5.93
Comparison happiness: 5.95
Why comparison is happier than reference:



This word shift shows why the ninth decile of a palliative care conversation is happier than the eighth decile.

Example shift using LabMT

Reference happiness: 5.95
Comparison happiness: 6.08
Why comparison is happier than reference:



This word shift shows why the tenth decile of a palliative care conversation happier than the ninth decile.