Conversational Modeling from a Process-Oriented Perspective

Tientso Ning

Fall 2020

1 Abstract

Conversational modeling has made significant strides to allowing computers to better understand Natural Language. However, current works in this field focus more on a quantitative approach based in modeling the relationship between words and concepts, utilizing those models to establish understanding of intent. This quantitatively influenced approach generates consequently frustrating tools for language learners. We present an alternative expansion for an organizational structure based on a more Process-oriented method to language modeling, and set the groundwork for further creation of language learning tools that aim to avoid the current limitations.

2 Acknowledgements

Thank you to the wonderful modeling team led by Didier Buchs, as well as to the rest of the assistants, Dimitri Racordon, Damien Morard, and Aurelion Coet. Your support, guidance, as well as trust in me, made this entire process educational, rewarding, and an enriching experience.

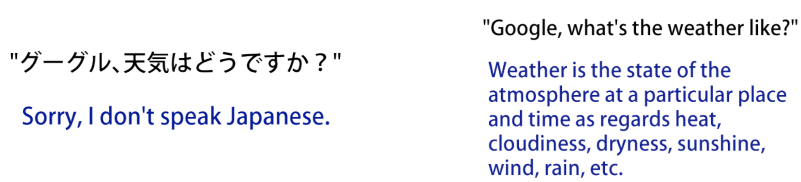
Thank you to Kense for keeping me on task, inspiring me to pursue this direction, allowing me to believe in myself, as well as for everything you've done to make all of this possible.

Finally, *"Kärt barn har många namn."* A dear child has many names. To all those who adorned me — the honor of being dear — Thank you.

3 Introduction

Learning a second language, specifically known as L2 acquisition, is a process in language learning that fields tons of research. And with technology becoming more and more an integral part of society, and with the reach of the internet expanding, the intersection of Language Learning and technology also becomes more relevant as a result. Learners growing up in this "Wi-Fi soup" have at their constant disposal potentially new tools and new methods to engage in learning a new language. It is thus important to look at those tools, and explore how to utilize technology to improve the processes of Language Learning in an effective manner.

We frame this discussion of the intersections of language learning and technology through the lens of the “Language Problem.” It is the problem of being able to simultaneously understand the quantitative aspects of language (i.e: word choice, word-order, etc), as well as the intent of language (i.e: tone, subtext, inference). Both are important to successfully understanding and correctly using a language, and as a result both are important when it comes to learning a language. To illustrate the challenge of solving this problem, we present the two aspects of the Language Problem through the example in Figure 1.



(a) The Quantitative Aspect.

(b) The Intent Aspect.

Figure 1: The two aspects of the Language Problem

To reflect the quantitative aspects of the Language Problem, we illustrate the case where a speaker does not understand the words chosen (in Japanese: ”Google, how’s the weather?”). Although the speaker may be able to guess the intent from recognizing the phrase as a question (from the intonation, and extrapolating knowledge from their own language to apply to this situation), the speaker will not be able to understand the meaning because the choice of words is unfamiliar to them and as a result, can choose words that creates an incorrect response (in this case, the question is not ”Do you speak Japanese?”). Here the quantitative aspect gates the intent aspect, as the inability to pick out the correct words negatively impacts the understanding of intent. While in the second case, the disconnect is in understanding the intent of asking about the weather alone. One can argue that the responder chooses the correct words, but delivers on a different intent entirely (describing what the weather is like as a concept, not about the current state of the weather). This case being the pre-eminently studied problem for machines [14]. The challenge for human beings when it comes to the Language Problem is fundamentally similar, although executionally different when compared to the problem for machines. Both aspects are possible and common mistakes for humans and machines alike. There are exceptionally cases where one can have no understanding of the quantitative, but correctly assess intent (i.e: ordering a drink at a foreign cafe), and vice-versa (i.e: Hearing your name in the middle of a foreign sentence). As a language learner, it can be seen why even in the extreme-cases, simply isolating one aspect when in the process of learning is insufficient, for both humans and machines.

When discussing the interactions of technology and language, most look no further than the term "Natural Language Processing," and while this branch of computer science accurately depicts addressing an aspect of the Language Problem and working with how machinery can understand human language [10], there exists many assumptions and perspectives that result from the common approaches in this field, creating a situation where one misses the forest for the trees. One can argue that the quantitative methods leak into the mindset of how one can approach solving the intent aspect of the Language Problem, poisoning the approach. Although one can simulate correct responses to intents through quantitative approaches in deep-learning, allowing a computer to give accurate responses, it is clear that computers do not capture the concept of the words the same way humans do [14]. The tools that are created from these approaches thus suffer from failing the intent aspect of the problem. We argue that current trends in Natural Language Processing focus on the more quantitative aspect when dealing with language, such as using words as the basic unit, as seen with classical machine learning and even more so with deep learning models. Relationships between words, intent, etc, are quantifiable values based on frequency counts derived from data. As a result the tools resulting from a successful application of current quantitative methods tend to find themselves to be tools that in function serve to replace the need for human language learning. In the logical conclusion of Neural Machine Translation, the "Language Problem" would be considered solved by the machine being the perfect universal translator. The human never needs to speak more than one language to resolve the issue of communicating. It would be analogous to creating a robot in basketball to throw the perfect free-throws, rather than creating a tool to help perform post-shot analysis, if the end-goal was to decrease missed free-throw percentages. The situation quickly devolves to methods of teaching that seem to treat human language learners as machines.

Perhaps inspiration can be taken from another field that works more exclusively with processes in less quantitative ways. The field of Process Mining works with representing data in this manner. However, it is to be seen whether these techniques can be applied to the context of a conversation, which differ in makeup when compared to traditional processes seen in Process Mining, which are based primarily in actions. Thus our aim in this paper is to identify and evaluate current methods in modeling language and processes, in order to find an intersection for which a process-oriented conversational model can be built upon.

3.0.1 Research Questions

Our main questions center around the approaches involved in creating tools for Language Learning, and what part of the Language Problem they address and are influenced by.

- How can we represent conversational data differently in order to consequently apply a less quantitative approach to modeling?
- What Process-Oriented methods that are currently being used in other fields can we apply to our models which work primarily on conversation?
- Can these methods lead to creation of tools that aid Language Learners in new ways?

3.0.2 Contributions

Our extension to works in similar fields serves as a starting point for an approach to language modeling that aims to create better analysis tools for language learning.

In the following section, we will present a State of the Art, detailing current approaches in the field which address Language Technology, as well as tangential fields which utilize methods we are interested in.

We then provide in Section 5 the details to relevant definitions and formalities necessary.

Finally in Section 6, we detail the organizational setup of our data and the methods applied to expand conversational modeling, in order to provide a more Process-Oriented approach to addressing the Language Problem.

4 State of the Art

4.1 Natural Language Processing

Natural Language Processing (NLP) aims to solve our defined "Language Problem" for machines. NLP methodology has seen three main periods of change, including the current period. First is Symbolic NLP, which were largely rule-based approaches using a previously defined set of linguistic definitions and constraints that were then applied to data in order to emulate language understanding. This group of rules, called Context Free Grammars, of course was unable to solve the problem as human use of language is far more complex than just a strict set of rules [10]. The second period is Statistical NLP, which more closely resembles the current period, and is what current NLP work is mostly based on. Statistical NLP is the quintessential quantitative approach to solving the Language Problem. The main boom of Statistical NLP came with the advent of machine learning, which used forms of supervised learning to leverage increasing volumes of text data. In essence, instead of a rules-based approach, NLP shifted to a numbers-based approach to processing language [12]. The third and current period of NLP is the deep-learning era, which is a further extension of Statistical NLP that leverages higher volumes of data, as well as more complex model setups [12].

Current leading NLP methodologies for processing language revolve around utilizing Word Embeddings, which are matrices that represent the relationships of words to every other word in the vocabulary, denoted by a numerical value. The creation of this embedding is thus crucial to the success and performance of the model in processing and understanding text, and can also be learned by machine learning models, resulting in what is called representational learning. One of the most popular (and widely used as a standard benchmark) word embedding system is Word2Vec [13], and it serves as a good baseline for developing more advanced models.

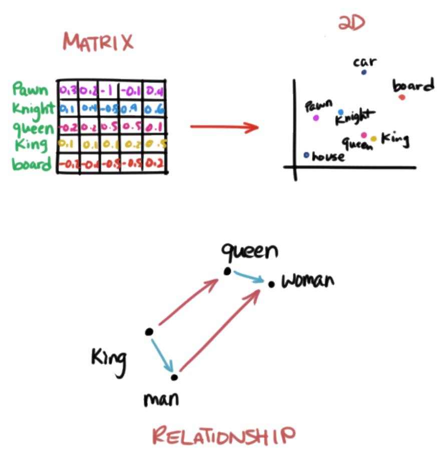
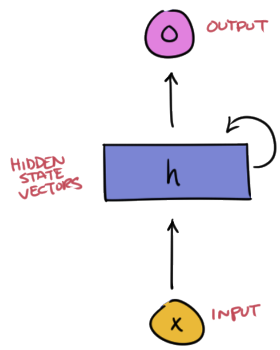


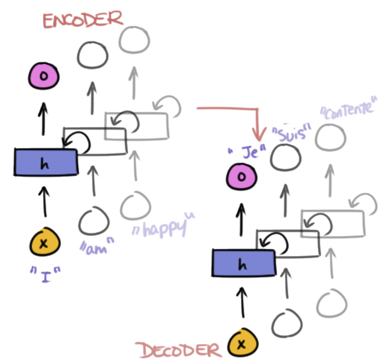
Figure 2: Word Embeddings in its different forms.

The visualization provided in Figure 2 displays common graphical representations of the word-embeddings concept. Matrices store large volumes of numerical data, comparing the relationship between each word in the vocabulary. This relationship can be visualized in a 2D-axis to show visual distance, and those relationships often carry properties humans can intuit (such as King is to Man, as Queen is to Woman).

Building on top of word embeddings, machine learning models can then learn to translate between two languages, called machine translation (or neural machine translation, NMT), with one of the popular architectures being Sequence-to-Sequence (Seq2Seq) [4]. The Seq2Seq architecture at the most basic level uses a Recurrent Neural Network (RNN) set up, leveraging an Encoder-Decoder model in order to iteratively produce translations. The quantitative aspect of the Seq2Seq architecture lies in its iterative calculations made for each token. The encoder takes each token and the context in order to produce what is called a hidden-state vector, which is a matrix of numerical data containing the relationship of words that the model has learned at each step. The decoder then takes this in order to produce and output a sequence of words in the desired language, completing translation.



(a) The Recurrent Neural Net.



(b) The Encoder-Decoder architecture.

As seen in the visualization above, the Encoder-Decoder system can thus be understood as two Recurrent Neural Nets working in concert to produce a translation, for instance, taking the source language (English) to the target language (French).

More advanced models in NLP today apply the leading structure called Adversarial Learning [3]. The structure utilizes two designated models each responsible for part of a min-max optimization problem, and sees good performance for NMT tasks, as they help to make the Decoder less dependent on the output of the Encoder and thus creates a more flexible translation system [2].

With developments in Machine Learning that address varying access to training data, theoretical explorations of new systems will always find a way to be utilized in NLP, improving the quantitative methods to generate better and better translations [12][10].

4.2 Language-Learning Applications

Language Learning applications in the market today can be categorized into four main categories. Vocabulary Trainers are the most popular and widely-known category, which focus on teaching language through a focus on foreign vocabulary, and aim to retain learners' attention through gamifying the experience or providing iteration on difficulty, similar to vocabulary quizzes in a traditional classroom experience. The primary method of retaining learners' interaction with the application is through notifications, as most of these applications are mobile. Examples of popular vocabulary trainer applications are DuoLingo[1],

[1]https://www.duolingo.com/approach

Memrise[2], Drops[3]. These vocabulary trainers provide easy accessibility through a freemium model, but lack any interactive exchange in the target language with human peers. The second category are Course-Based applications, which focus on being a virtual port to the traditional physical classroom experience. Examples of these applications are Babbel[4], Rosetta Stone[5], Pimsleur[6]. These applications for the most part provide content created by language teachers, and are most akin to a more engaging textbook. They are rather inflexible to differing learning styles, but could provide room for interactive exchange in the target language with human peers if used in a classroom setting. The third category is the Content Aggregator applications, which apply machine learning models on student learning data in order to either provide appropriately-leveled media or text content, or customize learning plans. This category of applications lacks interactive exchange with other human peers, but the process of consuming media is often cited as one of the preferred methods of learning a new language, and thus the use of machine learning brings a unique opportunity to applications in this category. Examples include LingQ[7], Fluent U[8]. The last category is the Language Exchange applications, which are often in the form of forums or a social media structure, providing an avenue for users to exchange social interaction in many languages. Users can interact with each other, displaying their proficiency in listed languages, assisting each other with voice recordings or corrections. Examples of applications in this category include HelloTalk[9], LanguageHero[10], Tandem[11].

4.3 Process Mining

Process Mining is the task of analyzing processes, which are sequences of defined actions, for the purpose of identifying trends, patterns, and points of interest. This is done through a process that involves three main parts. The first is Discovery. In Process Discovery, algorithms are applied over a large collection of processes, called the event log. The goal of discovery is to create a model that represents the activities seen in the processes of the event log[15]. The second part of Process Mining is Conformance Checking. This is the task of comparing the previously discovered model against the processes in the event log, in order to check for discrepancies and adherence to certain properties that may prove to be useful in analysis[15]. The third part is Performance Mining, which is concerned with how the model suggests performance aligns with certain variables such as time, cost, etc. The goal is to improve the performance of the

[2]https://www.memrise.com/
[3]https://languagedrops.com/
[4]https://www.babbel.com/
[5]https://www.rosettastone.com/
[6]https://www.pimsleur.com/
[7]https://www.lingq.com/en/
[8]https://www.fluentu.com/en/
[9]https://www.hellotalk.com/
[10]https://mylanguagehero.com/
[11]https://www.tandem.net/

theoretical processes with the model, such as increasing efficiency or identifying and removing bottlenecks[15].

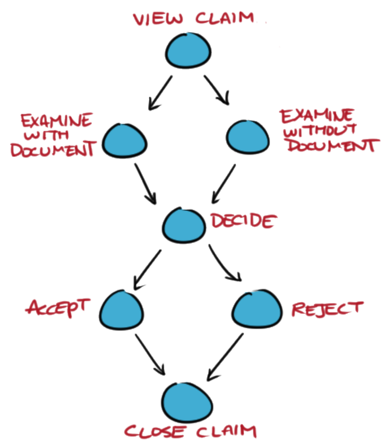


Figure 4: A sample generic process of dealing with a claim.

As seen in Figure 4, this sample process can be a result of automatic data collection on a service that involves dealing with claims (i.e: an airport service). The process involves actions such as examining the document, as well as deciding to accept or reject the claim before ultimately closing it. Many of these processes that involve different actions and aspects of the system can be collected into an eventlog and be used for Process Discovery.

One of the baseline algorithms used in the Process Discovery phase of Process Modeling is the Inductive Miner, which uses recursion strategies to guarantee a good process model[15]. These algorithms and many variations of such have seen applications to business processes, as well as educational processes focused on student learning strategies, and many healthcare applications.

Many business process mining applications focus on predictive process modeling, which involves applying process mining techniques to analyzing trends in order to inform future trends [9]. These predictive modeling techniques often intersect with machine learning and deep learning models.

On the healthcare side, process mining developments focus on the issue of privacy-preserving techniques, in order to make use of process mining while maintaining and ensuring patient confidentiality [6][7].

These techniques have not yet seen widespread usage in conversational modeling.

5 Definitions

In this section, we will detail some of the concepts, specifically the calculations and formulations, involved with the implementation of our process-oriented approach. These formulations act as the structure and framework for the theoretical approaches that follow. We start with the details of the conversational data itself, and proceed with different methods of evaluation performed on those data.

5.1 Definition of Data Terms

Conversation data is obtained from the Open Subtitles website[12] (as SRT files). Each file contains speech (and occasionally actions), separated line by line, and will be referred to as an utterance. The labelset L is defined as a set of labels, pertaining to the classification of utterances.

closed.question	Question that can be answered with a binary response.
open.question	Question that requires more than a binary response.
respond.agree	Any response that answers the given question, or a statement in the affirmative.
respond.deny	Any response that denies answering the given question, or a statement in the negative.
deflection	Any response that does not clearly answer or deny answering the given question.
give.statement	Declaring a fact or general statement.
give.opinion	Declaring an opinion.
recall	Giving information from one's past.
use.social.convention	Greetings, social niceties, etc.
relax.atmosphere	Jokes or statements meant to change the mood.
quoted.speech	Utterances that include direct quotation or imitation.
inner.dialogue	Utterances that are not said aloud, narration or inner thoughts.
exclamation	An utterance that expresses strong emotion, i.e: shouting, battlecry, etc.
misc	Any non-verbal utterance such as gestures, or singing.

Table 1: The labelset used.

In Table 1 is the full labelset used for our organizational setup. These four-

[12]https://www.opensubtitles.org/en/search/subs

teen labels were chosen to represent common conversational intentions, which dictate the direction and eventual purpose of a conversational exchange. Selection of labels directly impacts the results (as explained in Section 6), and these fourteen labels were chosen to maintain a broad and generic labelset over the conversational traces we utilized in our eventlog, in order to serve as a starting point. More specific and nuanced labels could be chosen as an exploration for further works, as well as a different select number of traces for the eventlog. (For more details on the effect of data selection on the results, refer to sections 6.1 and 6.4).

Each conversation file is defined as a trace, S, which contains a sequence of utterances $(u_1, \ldots, u_n)$. Each utterance has a corresponding label $(l_1, \ldots, l_n)$.

"All right, let's do this one last time."	exclamation
"My name is Peter Parker."	use.social.convention
"I was bitten by a radioactive spider."	recall
"And for 10 years"	recall
"I've been the one and only."	statement
"Spider-Man."	statement
"I'm pretty sure you know the rest."	give.opinion

Table 2: Sample Trace and corresponding Labels, from the film "Spider-Man: Into the Spider-Verse"

In Table 2, a snippet of the trace from the film "Spider-man: Into the Spider-Verse" is shown, with the utterances on the left column, and their corresponding labels in the right column.

For convenience, calculations involve only the labels of traces, thus the terms "sequence" (of corresponding labels) and "trace" will be used interchangeably to refer to the trace. A collection of traces is defined as an Eventlog, $E = (S_1, \ldots, S_M)$, where M is the total number of traces collected.

5.2 Definition of Distance

The traditional definition of distance is a numerical measurement of space between two points. In Natural Language Processing, the notion of distance is abstracted and used to define how different two words or documents are. The further the distance, the more different the two concepts are. In a Process Mining perspective, the notion of distance is abstracted and used to define how different two traces are. The further the distance, the more different the traces.

For our purposes, we compare distances between sequences of labels, which are numerical values to define how different two traces are from one another. These distances are derived from different calculations and methods, generally involving frequency and specific location of occurrence. We will outline a couple of common approaches.

There are four common methods in frequency-based analysis: Bag of Words, K-grams, Hamming Distance, and Edit Distance (Levenshtein Distance). The

Bag of Words method defines a vocabulary of words, which in our context is equivalent to our labels and labelset. The method takes a trace S_i and counts the appearance of each label in the labelset for the entire trace.

"Wow. Is it too crazy?"	closed.question
"No, man."	respond.deny
"Miles, I see exactly what you're doing there."	give.statement
"Yeah."	respond.agree
"You know, me and your dad used to do this back in the day."	recall
"Stop lyin'. It's true."	respond.deny
"Then he took on the cop thing, and I don't know"	respond.agree
"He's a good guy, just"	give.opinion
"You know what I'm sayin'."	give.statement
"All right, come on, man. I gotta roll."	give.statement

Table 3: Sample Trace, from "Spider-Man: Into the Spider-Verse."

Label	Frequency
closed.question	1
respond.deny	2
respond.agree	2
give.statement	3
recall	1
give.opinion	1

Table 4: Bag of Words for the sample trace.

In visualization above, the Bag of Words method is shown over the sample trace of labels. The frequency is given for the provided labels in Table 4 as they occur in the sample trace from Table 3.

The K-grams method is a variation of the Bag of Words method, where each word is defined as a k-gram. In context, a k-gram is a sub-sequence of labels, where k defines the length of the sub-sequence.

Figure 5: A count of 2-grams

In the visualization provided in Figure 5, a k-gram of two is used on the sample sequence "ababdacaaab", for which every occurrence of two labels as a pair is counted.

The Hamming-Distance method is a metric that counts points of difference between two given sequences. Given two sequences S_a and S_b, the hamming-distance h_d is the total count of differences in each utterance between the two sequences, where $0 < h_d < n$.

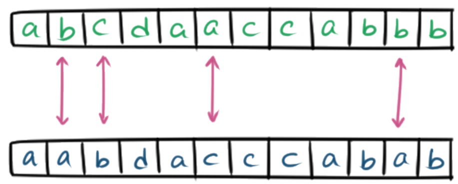


Figure 6: A hamming distance of 4.

In Figure 6, two sequences (green and blue) are shown with four labels being different. This represents a hamming distance of four, as the two sequences differ only in those four locations.

The Edit-Distance method is a metric that counts the minimum number of operations (from insertion, deletion, substitution) required to turn one sequence into another. For generic edit-distance (also called the Levenshtein Distance), all operations have the same uniform cost (of one). Edit-Distance variations will often assign different costs for each operation.

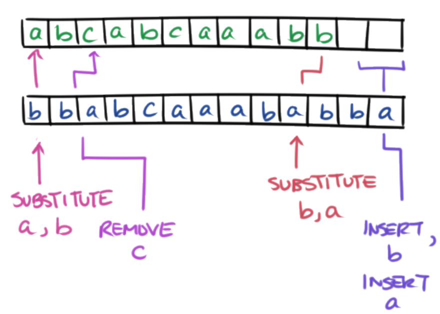


Figure 7: A Levenshtein Edit Distance of 5, showing substitution, insertion, and deletion needed to transform the top trace into the bottom trace.

In Figure 7, two sequences are provided (green and blue) of differing lengths, and the required operations needed to transform the top sequence (in green) to the bottom sequence (in blue) is shown. This pair of sequences will have a Levenshtein Distance of five, as five operations (substitution, removal, substitution, insertion, insertion) are required to transform the top sequence to be identical to the bottom sequence. Since the variation is the Levenshtein Distance, all operations have a cost of one, which results in the final distance remaining as five.

5.3 Definition of LTL

Linear Temporal Logic (LTL) is an extension of formal logic that adds the modalities to express time-related conditions. LTL is made up of propositional variables AP (true, false), logical operators (not, or, and), and the temporal modal operators (X, U). The grammar of LTL can be defined as follows:

$$\phi \coloneqq \alpha \mid \phi \wedge \phi \mid \phi \vee \phi \mid X\phi \mid \phi U \psi$$

where α indicates the current label, with $\wedge$, $\vee$ being conjunction and disjunction, $X\phi$ indicates that ϕ occurs as the next label, and $\phi U \psi$ indicates that ϕ holds true until ψ occurs sometime in the future. Both ϕ and ψ represent sets of labels.

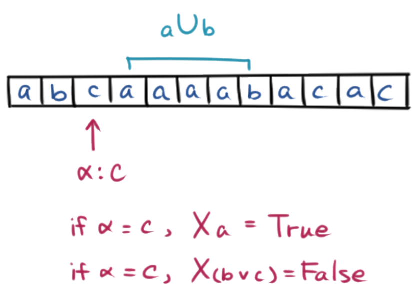


Figure 8: LTL applied over the sample sequence, demonstrating the X operator, and the U operator.

In Figure 8, a sample trace is shown with the current position over the third label in the trace (label "c"). For the current position, the evaluation of the LTL statement of the next label being "a" (denoted Xa) is True, while the evaluation of the next label being either "b" OR "c" (denoted $X(b \vee c)$) is False. Additionally, a sub-sequence starting from label 4 to label 8, is shown as a pattern that satisfies the LTL statement of a sub-sequence starting with the label "a" and ending with a label "b" (denoted aUb, read "a" until "b").

5.4 TF-IDF

Term Frequency, Inverse Document Frequency, is a calculation from language processing that deals with determining importance of terms appearing in vocabularies. It is made up of two parts, Term Frequency, and Inverse Document Frequency.

$$\text{TF-IDF} = \text{TF} * \text{IDF}$$

Each part is responsible for a separate value that is of importance to ranking terms. A term is defined as a word in the vocabulary, and a document is defined as a collection of words. Term Frequency is calculated by taking the numerical count of appearance of the specific term t_i in a given document d_j.

$$TF = count(x),\ x \in L$$

Term Frequency represents the importance of a term based on how often it appears over all documents.

Inverse Document Frequency is calculated by taking the log ratio of the total number of documents over the given term's number of document appearances.

$IDF = log\frac{N}{doc(x)}$, where N is the number of documents, and $doc(x)$ is the number of documents term x appears in.

Inverse Document Frequency represents the importance of a term by penalizing terms that occur over all documents, and is the part of the calculation that accounts for common stop words (the, as, a, etc).

Document	Text
1	"Today's such a gorgeous day, don't you think? Guess I should put some sunscreen on."
2	"Hey hurry it up, some of us have withdrawals to make you know?! Put the money in the bag! In the bag!"
3	"I guess that makes me, Oscar, the millionaire. The millionaire. The millionaire."
4	"There have always been ghosts in machines. Someday they'll have secrets. Someday they'll have dreams."
5	"Hey, how do I say 'Please understand my situation' in Korean? Why are you laughing?"

Table 5: Example Documents for TF-IDF calculations.

Term	Frequency	Document Frequency
Today	1	1
Guess	2	2
They	2	1
Hey	2	2

Table 6: Sample term frequency and document frequency count for words.

A short sample term and document frequency count is given in Table 6, counting from selected words occurring in Table 5. We can see that words that appear across in multiple documents (such as Guess or Hey) would appear to be more relevant to analyze, while words that appear multiple times but only in a singular document (such as They) may have less relevance to the other documents surrounding it.

5.5 PF-ITF

Pattern Frequency, Inverse Trace Frequency is the application of TF-IDF, where the smallest defined unit "term" is replaced with a bigger defined unit "pattern." It follows the same fundamental principles as TF-IDF. A pattern p is defined as a sub-sequence of labels u.

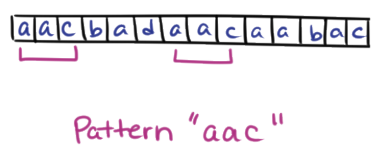


Figure 9: Sample pattern "aac" for the sample trace "aacbadaacaabac".

Figure 9 shows the two times in the sample sequence "aacbadaacaabac" where the pattern "aac" occurs. Calculations for PF-ITF is done where pattern frequency is counted, and the inverse trace frequency is how many traces contains this pattern in the eventlog.

6 Process-Oriented Conversational Modeling

In this section we will detail the organizational setup of the conversational data, which functions as the basis for the techniques that follow, as well as the theoretical and technical implementations of those techniques in order to demonstrate a process-oriented approach to our tasks in language modeling. We present some of the motivations as well as the considerations taken when applying these techniques to the conversational data, as well as present the deviations away from quantitative approaches.

6.1 Data Collection and Setup

One of the main difficulties in NLP is the task of data collection and selection. Due to the large volume of data required to achieve high performance, in addition to the quality level of said data, the whole operation is both costly and critical to the success of any and all NLP tasks. This is no exception in our case. There is no argument against the importance of this data, and careful consideration is to be paid to the data chosen. However, this being stated, the method in which we then process the selected data, such as choosing features, removing stop-words, structural organization, etc, are up for discussion in terms of useful variations, and we will present our choices as well as motivations in this section.

We chose conversational data to be the center of focus, as our goal was to explore a more process-oriented approach to modeling conversation. There were a variety of genres within what qualifies as conversational data, ranging from interview-based talk shows, multiple-account documentaries, to scripted dialogue. Several of these were chosen from the Open Subtitles website and processed in accordance to our outline in section 5.1. Additionally, the full list of traces and their identifier number are listed in Table 8. The organization we chose was based on the notions of treating conversation exchange as one would treat a process: a sequence of actions which entails additional resulting actions, from the start to the end of an individual process. In this way, each utterance was considered as an action, and appropriately labeled based on the intent of the utterance. The generation of the labelset itself is also of significant importance, as the classification of allowed actions dictates the appearance of the resulting process. We chose to generate the labels to be centered around generic actions and classifications seen in conversation. Note that while the choice of the labelset and its specificity in terms of classification (i.e: the detail and nuance in minutia between two given labels) affects the resulting processes, our labelset was chosen to be broad in order to set a good starting-point, rather than start with a more strictly defined labelset. The full labelset can be seen and referenced back in Table 1.

"Oppa, do you remember"	open.question
"how we had to record that whistling sound 1,000 times?"	closed.question
"That I'll never forget"	respond.agree
"I cannot forget that"	respond.agree
"And we almost didn't come out with "Whistle" for our debut song, but..."	give.opinion
"Too many people thought it was risky."	give.opinion
"But last minute, we were like, "It has to be that one.""	recall
"'Cause it has this weird country vibe to it,"	give.opinion
"and it's super minimal."	give.opinion
"It just sounded too empty for a lot of people."	give.opinion
"Most people were against it. We just pushed forward, whatever."	recall
"Hey, you know that photo there?"	open.question
"That's our very first photo session as BLACKPINK."	give.statement
"Yeah."	respond.agree
"It's oppa's favorite photo."	give.opinion
"As soon as I saw that photo, how it turned out, I was like,"	recall
"Oh, this is gonna be fun."	recall

Table 7: Sample Trace and corresponding Labels, from the film "Blackpink Light Up the Sky"

In the partial sample trace provided in Table 7, we can see a short conversation segment involving multiple speakers. It is important to note that the utterances are partially broken up sometimes even within the same utterance by a single speaker (i.e: "Oppa, do you remember" and "how we had to record that whistling sound 1,000 times?"). This leads to an interesting effect in the labeling of traces, as the intent of the utterance can suddenly shift mid-utterance, for instance going from an open question to a closed one. We argue that this organizational setup, given a well-defined labelset, can have the ability to capture intricacies in conversation, even when looking at one user's utterances. This setup also contributes to chains and patterns discoverable using methods we detail in later sections (i.e: In this segment, there is a chain of three "give.opinion" labels for one complete utterance talking about the "weird country vibe"). Consistency or variation in a single utterance is important to model, as it dictates the resulting "flow" of the conversation. The conversation data chosen were thus organized in the manner as shown.

There are a couple of benefits to the structural organization we chose for conversation modeling via our labelset that are important to outline. First, is the parallel to the notions of a process in an eventlog, bringing the perspective to modeling the "flow of conversation." If we use the intuition of quantitative methods in NLP as a reference, word embeddings utilize occurrence to establish context, which allows for a point of reference to understand and model relationships between words. In a similar intuition, we choose to identify actions (labels) which allows for a point of reference to understand and model this "flow" between exchanged utterances. We will forgo a formal definition of this "flow" of conversation in favor of an anecdotal example that will better present the concept. Early forms of chatbots that mimic conversation exchange struggled to maintain the illusion when confronted with situations where either the human input or chatbot output was not within the margins of expectation. This break in the "flow" contributes to the breaking of the illusion of conversation due to the fact that the output does not adhere to what we expect, and thus we question whether the entity on the other side is truly "human" (to which we already know the answer is no). In an exchange in a foreign language, this phenomenon also occurs, and while the resulting break in the "flow" can be attributed to different factors ranging from accent, lack of vocabulary, speed, etc, they yield a singular outcome: a breakdown of communication. It is also of interest to note that any one of the reasons for a break in the "flow" of conversation can overcome resiliency of any other factor. For instance, knowing the word for "beverage" in a foreign language, one can be in a situation when ordering food where the word for "beverage" is still the cause of the breaking of the "flow" due to speed or accent variation on that word. However, despite there being many potential causes for this disruption in "flow," the result is singular and thus can be accounted for when modeling. This approach to modeling and understanding failures not from the cause but from occurrence is the same intuition used in business contexts with Process Mining. The modeling never relies on knowing the root cause of the failure, simply that the failure occurs. Hence, the organizational setup allows us to examine and model conversations that can account for this "flow" in analysis and evaluation.

There is a notion of an "expected context" derived from "flow" and is important in conversation, and even more so for conversation in a foreign language. To use the chatbot example, if we are to believe that the other person exchanging conversation with us is "human," then that is the same to say that they have the burden of convincing us that they speak the same language, i.e: we must be able to confirm communication. In a given conversation, if there exists no notion of an "expected context" that one can anticipate or establish, then a conversation where both parties utilize their entire respective vocabularies but exchange non-sequiturs can be considered a fluent and successful conversation in a foreign language. This as we know, of course, is not the case in reality. Thus, modeling conversation in a process-oriented manner still requires a methodology for checking the behaviors of the exchange. This is possible with the label set we have established, and is another such benefit of a process-oriented setup, allowing for conversation to have naturally checkable "states." The hypothesis is thus, if we are able to model conversations in this manner, understanding points of failure and doing root-cause analysis during post-evaluation becomes more effective.

There are some challenges that arise, as well as additional considerations that must be had, given the organizational setup. One such major consideration is the labeling aspect of our data. There exists some level of abstraction to treating the conversation in a process-oriented manner, and this abstraction is applied when the label is applied to each utterance. But both determining the proper label for each utterance as well as establishing an entire labelset is a task that must be manually done, or done through some variation of machine learning. As a result, the setup of the data dictates that at some level, to a certain extent, there exists an assumption that two different sentences using different words can serve the same existential purpose in conversation. And although a potentially interesting concept, whether it is factually true or reflected in reality is not the primary concern of our work.

The abstraction and organizational setup of the conversational data collected can be referenced, and serves as the building block to additional notions important to modeling seen in later sections.

6.2 Distance of Processes

Distance is an important quantitative factor for establishing relationships between concepts. We mentioned and underscored the interest in expanding into less quantitative aspects, so why then is distance still so fundamental? Here is where we must take the time to reinforce the crux of our argument for a process-oriented perspective, that which is different from simply suggesting quantitative notions are useless. This is far from true. Differentiating different processes is still an important and necessary element in leading effective analysis and modeling. Comparing two traces in order to establish the difference requires this notion of distance. However, in a process-oriented setup, the comparisons made in order to calculate distance become slightly more complicated than our respective quantitative counterpart. Similar to word-embeddings where the notion of distance is established based on occurrence of each word in relation to another word, we can conceptualize processes in the same manner to establish distance. The variation comes in what additional factors are considered when applying a frequency-based algorithm to a process-oriented setup, which do not occur in the word-embeddings setting.

One such consideration is the mismatch of length between processes, which complicate the traditional distance concepts such as hamming distance. Additionally, order of appearance and contextual information adds a layer of complications to processes that we do not necessarily see in word-based distance concepts. For instance, basic notions of distance applied to two words do not differ based on which word appears first in the document, since in the majority of cases word orders do not alter the relationship between the two. However, this ordering is important and does alter the resulting implications of the process-oriented conversation. For instance, answering a question before a question is asked will imply massive difference when compared to a more expected answer-following-a-question ordering. Thus some variation on traditional notions of distance must be implemented to accommodate the nuances.

At the base of the distance metric, is the unit-cost edit distance algorithm, also called the Levenshtein distance. Since edit distance is based on the number of operations, it does not have the same restrictions when it comes to length when compared to the hamming distance. In order to obtain some notion of context for appearances of labels in the process, we follow the modifications made to edit-distance proposed in the works of [11]. The addition of substitution and insertion costs allow us to establish a relationship between the appearance of any label in any context, as given by the processes in our eventlog. This modification under our setup grants metrics to find separation between two conversational processes, provided an adequate eventlog is used.

Following the modifications, we can identify and separate our conversational traces even in our process-oriented setup. The full scores used to derive the distances are provided for Substitution and Insertion in the Appendix, under Table 14 and Table 15. The full distance derived are in Table 16, and the source of the traces are listed below in this section under Table 8.

Trace Number	Source
1	The Graham Norton Show s22e08
2	The Graham Norton Show s22e12
3	Black Pink Light Up the Sky
4	The Graham Norton Show s22e01
5	The Graham Norton Show s22e02
6	The Graham Norton Show s22e07
7	The Graham Norton Show s22e15
8	The Graham Norton Show s22e19
9	The Graham Norton Show s24e10
10	American Factory
11	Miss Americana Taylor Swift
12	Spiderman Into the Spider-verse

Table 8: Trace Number and their sources.

The twelve traces were used as the full eventlog to generate the substitution and insertion scores used to derive the distance values. The corresponding Trace Numbers are listed, and for simplicity are referenced as such (i.e: "Trace 11" refers to the trace for the film "Miss Americana Taylor Swift"). We will highlight a couple of notable results. However, it is worth mentioning that a distance metric established on a process-oriented conversational trace is able to provide a good sense of separation. This implies that a frequency-based algorithm, with shorter term dependencies (looking at co-occurrence and context) is still navigable on a process-oriented setup. Traditional metrics and important conclusions derived from these algorithms are thus compatible with the organizational structure (with some variation) and new derivatives based on this structure can be considered a proper expansion. Furthermore, the level of abstraction in treating conversational data as a process does not interfere with identifying aspects of the data enough to prevent separation, and suggests abstraction of conversation can be a favorable area for further exploration. In addition, just as machine learning has a concept of a semi-supervised model, a semi-abstracted conversational model may be able to leverage both traditional quantitative measures along with abstractions in order to increase the use-cases of conversational modeling.

To further provide warrants to this claim, we isolate some noteworthy values.

Substitution	Score
"respond.deny, deflection"	1.207
"respond.deny, respond.agree"	0.465
"recall, inner.dialogue"	1.773
"quoted.speech, recall"	-3.378

Table 9: Selected substitution scores.

In this segment of selected substitution scores, two labels are listed as the candidates for substitution and is given a score based on replacing the first label with the second. The score value can range from positive values to negative ones, with positive values indicating that the substitution is "favored," implying that the two labels share common co-occurrences and context in the eventlog, suggesting similarity in usage. Negative values indicate that substitution is "discouraged," implying that the two labels do not share common co-occurrences and context in the eventlog, suggesting similarity in their lack of connection. The value itself identifies to what extent the "favoring" and "discouragement" is reflected in the eventlog, with values closer to 0 being less strong, and bigger values being stronger. There are a couple of things to notice.

First is the entry "respond.deny, deflection" bearing a score of 1.207. This value suggests that the actions of denying to answer a question or answering a question in the negative has high contextual similarities to deflection, the act of not answering a question through vagueness or other means. This would suggest that conversational abstractions can be correlated even in a process-oriented setup, which reinforces the previous points made. Similar entries are "respond.deny, respond.agree" sharing a score of 0.465. Intuitively, one could argue that semantically, responding in the positive or the negative to a question makes a big difference. However, just as in quantitative aspects of NLP, classifying responses as an action within a conversation can be abstracted as similar notions despite having potentially opposite outcomes. Additionally, "recall, inner.dialogue" scores 1.773 displaying similar use-cases, while conceptual similarities such as "quoted.speech, recall" score a more drastic -3.378. This underscores the notion that these scores (and eventually distances) are still quantitative in nature, which are more heavily impacted by the volume of data provided. This also suggests that there is perhaps a level of abstraction that benefits from having a more specific labelset than the one we have chosen, as these nuances in concepts can be reflected and captured by such scores, and that leveraging small differences in modeling with larger amounts of data can provide even more insight on small nuances in conversation.

Insertion scores are also listed, and contain interesting conclusions themselves, we will omit them and present the more important similarity values to discuss the previous notion of separation from distance.

Trace	2	3	4	5	10	11	12
1	354.243	-60.860	786.543	-47.538	-494.732	-999.692	-1139.817
10	-637.114	-700.619	-449.934	-2277.230	x	-241.235	-421.178
11	-981.314	-939.975	-874.528	-1765.015	-241.235	x	-872.539

Table 10: Selected similarity scores between traces.

The values selected above in Table 10 are organized based on the similarity score derived from substitution and insertion scores modifying the edit-distance values, following the works of [11]. The trace numbers are listed as the horizontal and vertical margins, and the resulting similarity score between the two given traces are listed in the intersection. The similarity values can be interpreted in the same way as the scores, where a positive value implies correlation of co-occurrence and context, while a negative value implies the opposite, with the value itself being how strong the correlation is. Note that similarity values themselves can be converted to distance with an arbitrary equation, and the definition is not of great importance, for our purposes we will refer to similarity and distance interchangeably.

Trace 1 along with Trace 2 and Trace 3 display a good starting point for interpreting the similarity values. Trace 1 and Trace 2 are both from the Graham Norton Show, having a similarity value of 354.243 while Trace 3 can be seen having a negative value with Trace 1 of -60.860 which is to be expected as Trace 3 is a documentary on an entirely different subject. With some exceptions, notably Trace 5, it can be seen that Trace 1 shows closer similarity values to other traces of the same genre (i.e: the same show) than compared to traces from independent films (such as traces 10, 11, and 12). Additionally, these traces, although bearing a lower similarity score between them, still shows the similarity within "genre," as Trace 10 and 11 score -241.235 compared to against Trace 12, where the score jumps further to -421.178 and -872.539. Traces 10 and 11 are both documentary films involving interviews of icons in pop-music, while Trace 12 is a fully independent film, suggesting that these values can provide separation of these genres of conversation.

6.3 Pattern Matching and LTL

6.3.1 Pattern Matching

Pattern matching on long sequences to identify trends is one of the main aspects in Process Discovery, and the notion can be applied here to our conversational traces as well. The abstraction on conversational data allows us to identify patterns that define a "sub conversation" of sorts. A pattern emerging from a sub-sequence of labels would be useful to identify both for separation but also for post-exchange analysis. We isolate two main approaches to pattern matching for our conversational modeling structure. First is applying a pattern matching algorithm to find patterns that are recurring long-term dependencies throughout the sequence. Second is applying Linear Temporal Logic (LTL) expressions to check for properties within the trace.

Searching for recurring labels in a specific order of variable lengths is difficult when the patterns are not pre-established. This opens up a concern in terms of complexity as exhaustively searching leads to inefficient run times, which prevent real-time analysis applications. To avoid this, algorithms that take advantage of efficient searching methods must be employed. We take a variation from the works of [5], applying PrefixSpan to our conversational model. First, we are concerned with patterns within each individual trace rather than patterns that emerge as common between all traces in the eventlog. This variation allows for identifying sequential patterns inherent in a given conversation trace and is useful for analyzing behavioral trends specific to actions over a longer period. Once those patterns are identified, then looking for common patterns over the eventlog becomes possible.

The full results over chosen conversational traces are presented in the Appendix under Table 17. The rank score is derived from the PF-ITF variation listed in 5.5, and displays how strongly this pattern is reflected in the eventlog as a whole.

Pattern	Rank Score
give.statement, give.statement, give.statement, give.opinion	29.185661610679364
give.opinion, misc	28.476608295358854
exclamation, give.statement	26.254804992236714
closed.question, give.statement	24.5001135248475
misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc	24.469656581723036

Table 11: Selected patterns and their corresponding rank score.

We have selected a couple of patterns from the PrefixSpan algorithm implementation. The patterns are denoted by the labels in sequence of appearance, with their associated rank score. The higher the score, the more common as well as important the pattern is in terms of the eventlog provided. The PF-ITF calculation, based on the TF-IDF metric, works by evaluating the ratio of the appearance of the pattern in question versus the total number of traces that the pattern emerges in throughout the eventlog. This guarantees that patterns that score higher are ones that appear commonly in a given trace, but trend towards being unique to a few select traces rather than appearing in every single trace, which would indicate that they might have less explaining power in separating and being unique identifiers to a specific trace.

We can see that common patterns in conversation (especially in the context we have chosen) rank closer to the top, such as "give.statement, give.statement, give.statement, give.opinion" which is a pattern that occurs often in the talk-show traces, where the host prefaces his opinion with a couple of statements introducing the topic. We can also see that long chains of "misc" labels rank highly, as many (but not all) of these selected traces contain musical elements, which adds to the ranking of the "misc"-chains. We have omitted some of these chains from the full table for clarity. Patterns like "closed.question, give.statement" are also seen amongst our traces, where a straightforward question is asked, and in order to preface the question, is followed by a more objective statement. These patterns can be obtained from the collected traces in an eventlog at run-time and analyzed for root-cause analysis, as well as checking consistency of users' behaviors, for instance, looking at whether learners are utilizing common patterns in their dialogue exchange.

Additionally, these patterns can be used to weight individual labels in a trace in order to visualize a trace's distribution of patterns across the trace itself, and is helpful as an analytical tool to visualize the "flow" of a conversation as denoted by the patterns that emerge over the course of the conversation.

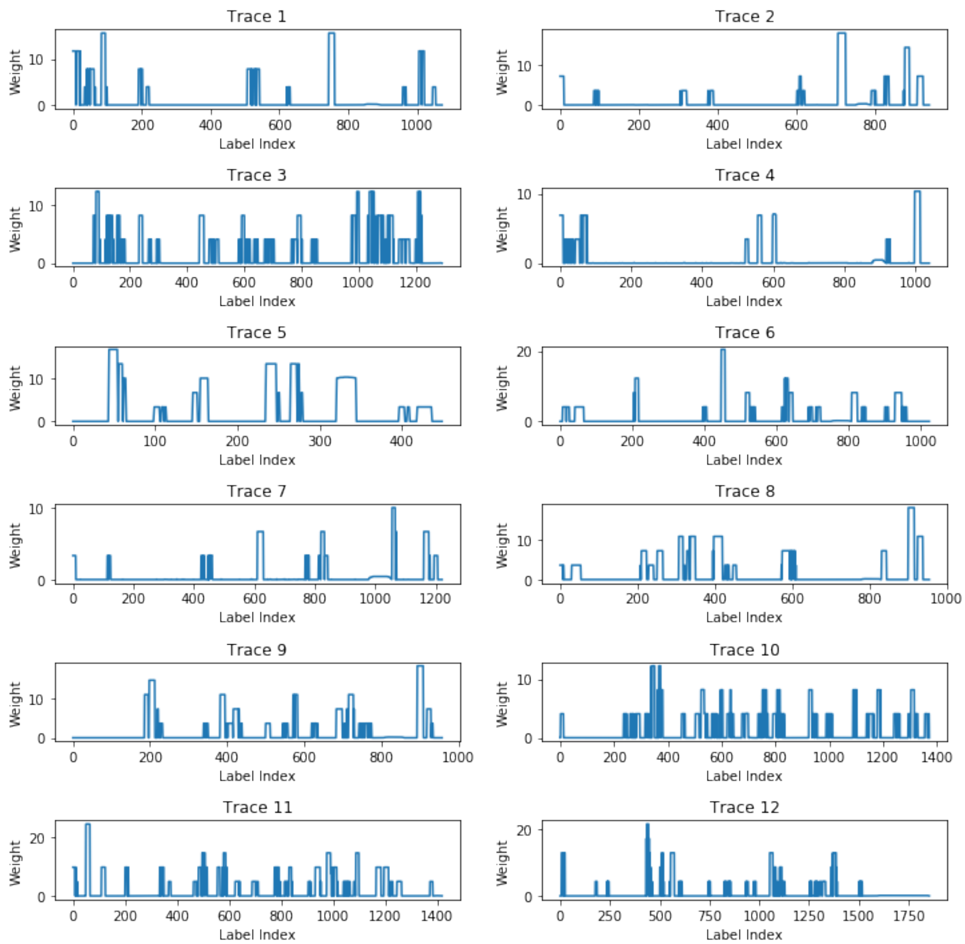


Figure 10: Weight visualization for all twelve traces.

Table 10 shows the twelve traces graphed to display the trace on the x-axis, with the y-axis being the weight of each individual label in the trace. This weight is determined by the rank score of patterns, where each given label in the sequence can be contributing to, or detracting from, a given pattern provided. For this table, we chose the Top-50 ranking score patterns. Selectively picking patterns can also yield visualization on specific behaviors, for instance, one could check the traces to see where large "misc"-chains occur (perhaps indicating musical sequences, like in our case).

It can be noticed that in our visualization, Trace 5 differs in the locations where the weights are most dense compared to Traces 1 and 2, which are of the same genre. This indicates that despite being the same genre, the patterns which emerge from this specific trace differs from the ones seen in Traces 1 and 2, which can explain why the similarity scores seen in the previous section varied as an exception. Traces 3, 10, 11, and 12 show similar density trends, and this serves as another example of these conversations being similar in function, as they were interview-based documentaries, with Trace 12 showing clear differences in the trend (as it is a normal film), with almost no patterns ranking highly near the end of the trace.

Isolating popular patterns and generating weights could prove to be useful in post-exchange analysis, evaluating whether certain labels or sequences of labels at certain points in a new conversation ”match” these expected popular patterns. For instance, post-exchange analysis can be done on a new-learner of a language, to evaluate whether their conversation meets the expected patterns of native speakers, which could point to difficulties due to unexpected contexts or a bigger problem in carrying conversation forward.

6.3.2 LTL

Applying LTL to conversational traces is quite possibly the biggest area for expansion that we get for our organizational structure. Traditional notions of model checking in the context of conversational data means that we can establish a “checkable state” inside any conversation. Since the conversational structure is abstracted, this concept of “checkable states” is arbitrary in definition but has useful analytical power still. For instance, in an intuitive sense, any given label in a conversation could be considered to be a “state” in the conversation, which would be arbitrary in the larger context of the trace. However, the existence of states in the conversational trace would then allow larger abstracted concepts to be checked or referenced, such as “flow” or even pre-defined rules such as “a question must occur before an answer is given,” “Between a question and answer, relaxing the atmosphere precedes any giving of opinions,” etc. Utilizing the extensive field already present in formal Modeling and Verification, LTL applied to conversational traces proves to be promising and worthwhile to explore as a tool for aiding language learning. Reference the sample patterns in Table 13.

Symbol	Label
p	closed.question
q	open.question
r	respond.agree
s	respond.deny
e	exclamation
c	recall
x	relax.atmosphere

Table 12: Label shorthand for Table 13

Scope	Pattern
Entire Trace	$\{p, q\}$ precedes $\{r, s\}$
Between $\{p, q\}\{r, s\}$	cost ≤ 5
Between $\{p, q\}\{r, s\}$	existence $\{x\}$ OR absence $\{s, e\}$
Between $\{c\}\{c\}$	absence $\{p, q\}$

Table 13: Sample patterns searched on given scope, defined by LTL.

Following the conventions in the works of [8], a scope and pattern are defined in order to search and verify different properties within a sequence. The Table 13 displays scopes on the left-hand side, with the following pattern that is searched for in a given sequence on the right-hand side. The scope determines the sub-sequence where the pattern is checked, returning a boolean value for whether the sub-sequence in question holds the constraints provided.

These four scope-patterns are simple demonstrations for constraints that can be checked to achieve a number of tasks such as: confirm that trace behavior is expected and valid (in cases of automatic trace collection), confirm trends exist in conversation as part of root-cause analysis (i.e: we suspect conversations break down often when between a question asked and an answer given, the number of actions exceed 5, or exclamations appear), find traces that adhere to certain requirements (i.e: determine which traces in the eventlog exhibit patterns deemed to be used often by native speakers).

Additionally, expansion on LTL-N as presented by [1] adds the ability to account for longer sequences that have more variability. This specific extension was chosen for its flexibility due to the variance of sequences often seen in conversational data. More complex expressions can be built or checked with LTL-N, and the limit allows for applications of cost functions, which are useful for optimization. Refer to Figure 11 for the visualization of the application of LTL patterns on a scope. Each scope can be checked against the original trace to see the label context. Checking for patterns within a defined scope allows for post-exchange analysis on different constraints, which can be used to improve performance as well as also assist in separation for classification.

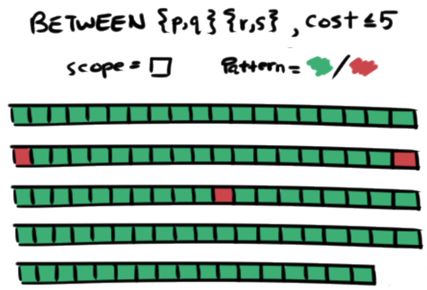


Figure 11: Visualization of all scopes found in a given trace, and whether the pattern is held True (green) or False (red).

In Figure 11, all 93 scopes, which are sub-sequences of the trace where a question (open or closed) is asked, and a definitive answer follows (agree or deny), are checked for whether the number of actions between the question-answer pair is less than or equal to five. As seen, there are three such scopes where this is violated, indicating that these three sub-sequences within the singular trace displays a long-term dependent pattern where in which a question is asked, and an answer is not given within the constrained period of five labels. This type of checking can be useful in root-cause analysis, to determine if an increase of these violations is correlated to break down of conversations, or perhaps any other defined pattern is common in one trace but not another, which can draw additional conclusions useful for learners.

6.4 Discussion

As presented in Section 3.0.1, we presented the following research questions, which we will discuss individually in more detail as follows.

- How can we represent conversational data differently in order to consequently apply a less quantitative approach to modeling?
- What Process-Oriented methods that are currently being used in other fields can we apply to our models which work primarily on conversation?
- Can these methods lead to creation of tools that aid Language Learners in new ways?

Representing conversational data in a format that allows for a consequently less quantitative approach to modeling naturally focuses less on the specific words (and their relationships between each word) and more on the known effects and intentions of those words in speech. Our choice of representing conversation data through the labelset is a function of such an intent. For our work the labelset was chosen to be generalist and a platform to build upon. This organizational structure that represents conversation as a process is critical to approach modeling in a more abstract manner. Even aspects of distance, which are quantitative in nature, see some level of abstraction when placed in context of traces (which represent conversational processes rather than entries of words, like a word-vector or a document). This structure also entails that resulting patterns and properties that are checked are more "behavioral" in nature, relating to the function they serve in the trace as a whole, rather than be representations of individual words. From an organizational standpoint this abstraction towards a conversational process reduces the likelihood of resulting models being translation tools or vocabulary-trainers, since the focus is no longer about capturing the relationship of words and concepts, but towards behavior and function in conversation. The specific choice of the labelset is of course open to variability. The selection of the labelset as mentioned in section 6 is specific to our domain and problem. Altering the labelset to address other domains (i.e: healthcare exchanges, business exchanges, etc) is an open question for further exploration. Additionally, a larger number of traces can be collected for analysis, as our work was limited to a manual selection of traces. Future works can look into automatic collection of data with some human oversight in order to bring in a larger collection of data for a more effective analysis.

As seen by the list of applications in section 4.2, the challenge is to present an approach to foster creation of tools which rely less on a mindset induced by quantitative measures. As we argued and demonstrated, process-oriented methods may better serve to reorient the mindset guiding the creation of tools from being engines that provide answers, expecting the human to self-correct like a machine learning algorithm, to being post-analysis tools capable of allowing learners to adapt to their personal needs. Current algorithms used in the Process-Mining fields focus on generating accurate Process models that are reflective of the system as a whole (from where the data is collected). These algorithms rely on capturing and understanding the relationship of the actions as they relate to the larger system they operate in. In a similar fashion, understanding the function of an utterance in the larger exchange of conversation (which those utterances "operate" in) seems a promising start to the creation of language learning tools. The effects of our work towards creating better language learning tools can be concentrated on two fronts. First is the effects of LTL and pattern matching, and the ability to check for properties or occurrences of labels of interest. The notion of property checking (via LTL) and pattern matching on labels rather than words broadens the applicability to tools focused on post-exchange analysis. Suddenly there is a way to check for valuable information after the exchange of a conversation for language learning students. Additionally, applications like weighting via pattern-matching (and the visualizations) can serve as a basis for not only comparing different conversations from different speakers, but also as a comparison between a singular speaker. This opens the avenue for creating tools that help and mark individual progress. It is important to note that none of these implementations remove the benefits provided by current language learning applications using quantitative frameworks. The two are not mutually-exclusive, which can only further the benefits experienced by language learners, as they now have additional tools at their disposal.

Additionally, algorithms and models developed for processes from other fields are suddenly free for the taking. Variations and adaptations can be freely made as there is at least some basis of comparison established for treating conversations as processes, to target different aspects of conversation. Process discovery becomes akin to finding new aspects of conversation, and different scopes of variability and specificity can be made to tackle different problems in modeling exchange of language. As the saying goes, what is old is new again. Lastly, new tools can be created to better supplement and accommodate new learners, which fit more with the goal of helping to learn a new language in a way that feels more human.

6.5 Future Works

There are three specific aspects of our work that we believe can bring additional research and contributions to the goal of bettering tools for language learners.

First, is the organizational structure. The labelset we chose was a broad, generalist, labelset intended to serve as a starting point and a basis for establishing basic relationships in conversations that involve language exchange. This labelset and its variability can be a topic of open research, as well as studying the effects of different structures accommodating different types of conversations. For instance, one may be interested to see how a more specific labelset geared at differentiating number of speakers can alter the makeup as well as the corresponding relationships established (such as distance, patterns, etc).

Second, visualizations from pattern-matched weighting on conversation traces opens an avenue for examining different types of conversations in comparison to one another, at a process-oriented, abstracted level. For instance, one could consider mining for patterns in a eventlog of student learners, based on an eventlog of native-speaker traces. Additional accommodations could be made to incorporate machine learning models for interests in separation, classification, or the like.

Third, creation of "checkable states" in conversations through LTL on conversational data represented in a process manner allows for more complex modeling on conversational flow. For instance, probabilistic models can be created based on different variables that can be verified by a formal modeling language. The resulting probabilistic models can see use similar to current work being done on probabilistic models in automation, bio-genetics, etc.

7 Conclusion

The future for language learning is bright with the amount of resources that are naturally created, as well as the number of technological contributions to the learning process both current and future. With exploration of both aspects of the Language Problem in mind towards the goal of improving language learning, creation of tools that allow learners to learn in a more human way seems all the more possible. The importance of approaches from a less quantitative and more process-oriented framework thus appears to be a promising start to a new side of language technology.

8 Appendix

Substitution	Score
respond.deny, give.opinion	-0.8144624951256775
respond.deny, misc	-4.787039500242717
respond.deny, recall	-3.759833903883851
respond.deny, relax.atmosphere	-1.1445663219573634
respond.deny, use.social.convention	-1.953942782461142
respond.deny, deflection	1.2073923927842225
respond.deny, respond.agree	0.4655414793076375
respond.deny, give.statement	-0.7610395082337496
respond.deny, inner.dialogue	-inf
respond.deny, closed.question	-0.6164065931496298
respond.deny, exclamation	-0.79070093151274
respond.deny, open.question	-0.7285767671280111
respond.deny, quoted.speech	-1.077670886384599
give.opinion, respond.deny	-0.8144624951256775
give.opinion, misc	-3.245577763395442
give.opinion, recall	-1.7728503581673456
give.opinion, relax.atmosphere	-1.4642122788712184
give.opinion, use.social.convention	-1.9447991272262384
give.opinion, deflection	-0.3724342657461309
give.opinion, respond.agree	-1.0852802030096529
give.opinion, give.statement	-1.0278190101071358
give.opinion, inner.dialogue	-3.5984601617817042
give.opinion, closed.question	-1.466547070115674
give.opinion, exclamation	-0.7815038855104308
give.opinion, open.question	-0.9591984330058558
give.opinion, quoted.speech	-0.5350067146143427
misc, respond.deny	-4.787039500242717
misc, give.opinion	-3.245577763395442
misc, recall	-3.5946827192875714

misc, relax.atmosphere	-3.0307266157577883
misc, use.social.convention	-3.493243736767782
misc, deflection	-4.0159904529046155
misc, respond.agree	-4.186599028574182
misc, give.statement	-4.26651113820859
misc, inner.dialogue	-4.296235652592653
misc, closed.question	-3.05669986160781
misc, exclamation	-3.1206114748820646
misc, open.question	-4.691036720907141
misc, quoted.speech	-4.696767145112473
recall, respond.deny	-3.759833903883851
recall, give.opinion	-1.7728503581673456
recall, misc	-3.5946827192875714
recall, relax.atmosphere	-1.7305104211915978
recall, use.social.convention	-1.7558897274998548
recall, deflection	-3.981299576017381
recall, respond.agree	-2.837530722318815
recall, give.statement	-2.867034143002071
recall, inner.dialogue	1.7733347675839626
recall, closed.question	-3.1380962534174865
recall, exclamation	-2.3223437074279514
recall, open.question	-1.5920217266546446
recall, quoted.speech	-3.378396688852107
relax.atmosphere, respond.deny	-1.1445663219573634
relax.atmosphere, give.opinion	-1.4642122788712184
relax.atmosphere, misc	-3.0307266157577883
relax.atmosphere, recall	-1.7305104211915978
relax.atmosphere, use.social.convention	-0.33690383034135607
relax.atmosphere, deflection	-1.195559503276419
relax.atmosphere, respond.agree	-0.9483744795190499
relax.atmosphere, give.statement	-0.8375590610009518
relax.atmosphere, inner.dialogue	-2.176463179108273
relax.atmosphere, closed.question	-0.7638322093641212
relax.atmosphere, exclamation	-1.0322068819441506
relax.atmosphere, open.question	-0.47339900658252193
relax.atmosphere, quoted.speech	-1.624612521058588
use.social.convention, respond.deny	-1.953942782461142
use.social.convention, give.opinion	-1.9447991272262384
use.social.convention, misc	-3.493243736767782
use.social.convention, recall	-1.7558897274998548
use.social.convention, relax.atmosphere	-0.33690383034135607
use.social.convention, deflection	-2.114044912189147
use.social.convention, respond.agree	-1.688865084036604

use.social.convention, give.statement	-1.159975836280691
use.social.convention, inner.dialogue	-1.5349743562940787
use.social.convention, closed.question	-1.5420487539380026
use.social.convention, exclamation	-0.9571206620863368
use.social.convention, open.question	-1.4881113722519201
use.social.convention, quoted.speech	-1.865016076611572
deflection, respond.deny	1.2073923927842225
deflection, give.opinion	-0.3724342657461309
deflection, misc	-4.0159904529046155
deflection, recall	-3.981299576017381
deflection, relax.atmosphere	-1.195559503276419
deflection, use.social.convention	-2.114044912189147
deflection, respond.agree	1.0290230359470363
deflection, give.statement	-1.5009074419713078
deflection, inner.dialogue	-inf
deflection, closed.question	-0.136952895720524
deflection, exclamation	-0.6923228960102885
deflection, open.question	-0.7659947241741337
deflection, quoted.speech	-1.1107639482251694
respond.agree, respond.deny	0.4655414793076375
respond.agree, give.opinion	-1.0852802030096529
respond.agree, misc	-4.186599028574182
respond.agree, recall	-2.837530722318815
respond.agree, relax.atmosphere	-0.9483744795190499
respond.agree, use.social.convention	-1.688865084036604
respond.agree, deflection	1.0290230359470363
respond.agree, give.statement	-0.6049638353113151
respond.agree, inner.dialogue	-3.5222679560247085
respond.agree, closed.question	-0.562149781535846
respond.agree, exclamation	-0.8439632032643855
respond.agree, open.question	-0.5356979298091689
respond.agree, quoted.speech	-1.1905283809746698
give.statement, respond.deny	-0.7610395082337496
give.statement, give.opinion	-1.0278190101071358
give.statement, misc	-4.26651113820859
give.statement, recall	-2.867034143002071
give.statement, relax.atmosphere	-0.8375590610009518
give.statement, use.social.convention	-1.159975836280691
give.statement, deflection	-1.5009074419713078
give.statement, respond.agree	-0.6049638353113151
give.statement, inner.dialogue	-3.918004071605737
give.statement, closed.question	-0.4955137269925848
give.statement, exclamation	-0.6542274972125894

give.statement, open.question	-0.9511725218575913
give.statement, quoted.speech	-0.4942535716524106
inner.dialogue, respond.deny	-inf
inner.dialogue, give.opinion	-3.5984601617817042
inner.dialogue, misc	-4.296235652592653
inner.dialogue, recall	1.7733347675839626
inner.dialogue, relax.atmosphere	-2.176463179108273
inner.dialogue, use.social.convention	-1.5349743562940787
inner.dialogue, deflection	-inf
inner.dialogue, respond.agree	-3.5222679560247085
inner.dialogue, give.statement	-3.918004071605737
inner.dialogue, closed.question	-3.8701577282506223
inner.dialogue, exclamation	-3.5130845985288714
inner.dialogue, open.question	-1.887533493765231
inner.dialogue, quoted.speech	-4.737272722356625
closed.question, respond.deny	-0.6164065931496298
closed.question, give.opinion	-1.466547070115674
closed.question, misc	-3.05669986160781
closed.question, recall	-3.1380962534174865
closed.question, relax.atmosphere	-0.7638322093641212
closed.question, use.social.convention	-1.5420487539380026
closed.question, deflection	-0.136952895720524
closed.question, respond.agree	-0.562149781535846
closed.question, give.statement	-0.4955137269925848
closed.question, inner.dialogue	-3.8701577282506223
closed.question, exclamation	-1.2050766441971537
closed.question, open.question	0.18778659043752988
closed.question, quoted.speech	-1.1513610557733367
exclamation, respond.deny	-0.79070093151274
exclamation, give.opinion	-0.7815038855104308
exclamation, misc	-3.1206114748820646
exclamation, recall	-2.3223437074279514
exclamation, relax.atmosphere	-1.0322068819441506
exclamation, use.social.convention	-0.9571206620863368
exclamation, deflection	-0.6923228960102885
exclamation, respond.agree	-0.8439632032643855
exclamation, give.statement	-0.6542274972125894
exclamation, inner.dialogue	-3.5130845985288714
exclamation, closed.question	-1.2050766441971537
exclamation, open.question	-1.3655019431742064
exclamation, quoted.speech	-1.232663132556828
open.question, respond.deny	-0.7285767671280111
open.question, give.opinion	-0.9591984330058558

open.question, misc	-4.691036720907141
open.question, recall	-1.5920217266546446
open.question, relax.atmosphere	-0.47339900658252193
open.question, use.social.convention	-1.4881113722519201
open.question, deflection	-0.7659947241741337
open.question, respond.agree	-0.5356979298091689
open.question, give.statement	-0.9511725218575913
open.question, inner.dialogue	-1.887533493765231
open.question, closed.question	0.18778659043752988
open.question, exclamation	-1.3655019431742064
open.question, quoted.speech	-1.3196644141920164
quoted.speech, respond.deny	-1.077670886384599
quoted.speech, give.opinion	-0.5350067146143427
quoted.speech, misc	-4.696767145112473
quoted.speech, recall	-3.378396688852107
quoted.speech, relax.atmosphere	-1.624612521058588
quoted.speech, use.social.convention	-1.865016076611572
quoted.speech, deflection	-1.1107639482251694
quoted.speech, respond.agree	-1.1905283809746698
quoted.speech, give.statement	-0.4942535716524106
quoted.speech, inner.dialogue	-4.737272722356625
quoted.speech, closed.question	-1.1513610557733367
quoted.speech, exclamation	-1.232663132556828
quoted.speech, open.question	-1.3196644141920164

Table 14: Substitution scores over all traces.

Insertion (Right/Left)	Score
respond.deny/respond.deny	-3.941447817644887
respond.deny/give.opinion	0.9407610990969307
respond.deny/misc	-1000
respond.deny/recall	-2.4378978958348365
respond.deny/relax.atmosphere	-1.6401545685615238
respond.deny/use.social.convention	-4.482086751617374
respond.deny/deflection	-1000
respond.deny/respond.agree	-3.190734075257016
respond.deny/give.statement	1.407123211642844
respond.deny/inner.dialogue	-1000
respond.deny/closed.question	1.6385263524526643
respond.deny/exclamation	-6.157902682789644
respond.deny/open.question	-6.15070718138544
respond.deny/quoted.speech	-6.746476437063179
give.opinion/respond.deny	-9.610732658820742

give.opinion/give.opinion	-0.9758753251957926
give.opinion/misc	-5.828661286751312
give.opinion/recall	-3.9614115972771566
give.opinion/relax.atmosphere	-4.976184897177227
give.opinion/use.social.convention	-5.090489350646208
give.opinion/deflection	-13.262262776480888
give.opinion/respond.agree	-5.4903839681462205
give.opinion/give.statement	-1.7150011615242398
give.opinion/inner.dialogue	-15.346031134173487
give.opinion/closed.question	-6.332010663858552
give.opinion/exclamation	-9.525297182314683
give.opinion/open.question	-11.21854139905157
give.opinion/quoted.speech	-10.72684781347897
misc/respond.deny	-13.332855874477053
misc/give.opinion	-4.339138642518244
misc/misc	-0.23208663685094358
misc/recall	-4.784911833308548
misc/relax.atmosphere	-6.472595333205477
misc/use.social.convention	-5.770207000037518
misc/deflection	-1000
misc/respond.agree	-8.31535559139428
misc/give.statement	-3.6623567503019583
misc/inner.dialogue	-13.625210853981068
misc/closed.question	-6.912359918386698
misc/exclamation	-10.76795102609715
misc/open.question	-13.667646120301464
misc/quoted.speech	-11.356524780370684
recall/respond.deny	-10.834168074374187
recall/give.opinion	-3.8404508424153794
recall/misc	-6.4880294258949505
recall/recall	-0.4938313797056882
recall/relax.atmosphere	-5.180358410570039
recall/use.social.convention	-5.536863766455648
recall/deflection	-14.686610885960329
recall/respond.agree	-5.991754497849507
recall/give.statement	-2.5129043910821918
recall/inner.dialogue	-15.033413649486722
recall/closed.question	-7.842515417087707
recall/exclamation	-11.591191320881647
recall/open.question	-14.490886415085962
recall/quoted.speech	-12.916730669321389
relax.atmosphere/respond.deny	-9.2190539299108
relax.atmosphere/give.opinion	-3.312799539202331

relax.atmosphere/misc	-5.815327789878618
relax.atmosphere/recall	-3.5534583000880224
relax.atmosphere/relax.atmosphere	-1.325240828928198
relax.atmosphere/use.social.convention	-3.4219398018342417
relax.atmosphere/deflection	-13.100065893693714
relax.atmosphere/respond.agree	-4.395931691994879
relax.atmosphere/give.statement	-1.0814287953084083
relax.atmosphere/inner.dialogue	-1000
relax.atmosphere/closed.question	-4.483922371499439
relax.atmosphere/exclamation	-8.741611922781237
relax.atmosphere/open.question	-10.319378922098188
relax.atmosphere/quoted.speech	-12.652113771942133
use.social.convention/respond.deny	-10.050363878205047
use.social.convention/give.opinion	-3.958082811762649
use.social.convention/misc	-5.252661854421851
use.social.convention/recall	-3.8880735295342923
use.social.convention/relax.atmosphere	-2.909366929366219
use.social.convention/use.social.convention	-0.9048914138006905
use.social.convention/deflection	-15.335766097067294
use.social.convention/respond.agree	-5.518290422292516
use.social.convention/give.statement	-1.5318678474750034
use.social.convention/inner.dialogue	-14.09760635987253
use.social.convention/closed.question	-5.607147845614609
use.social.convention/exclamation	-8.807387124712506
use.social.convention/open.question	-11.23315103058441
use.social.convention/quoted.speech	-11.980923379707198
deflection/respond.deny	-3.1699250014423126
deflection/give.opinion	1.9493231126003545
deflection/misc	-1000
deflection/recall	-1000
deflection/relax.atmosphere	-1000
deflection/use.social.convention	-1000
deflection/deflection	-3.2854022188622483
deflection/respond.agree	-1000
deflection/give.statement	2.1786460278454185
deflection/inner.dialogue	-1000
deflection/closed.question	4.250570954648042
deflection/exclamation	-1000
deflection/open.question	-3.5047152472667245
deflection/quoted.speech	-1000
respond.agree/respond.deny	-9.82117015833252
respond.agree/give.opinion	-1.9470345421263848
respond.agree/misc	-7.268580632385857

respond.agree/recall	-4.8581886178851725
respond.agree/relax.atmosphere	-3.9609096170609455
respond.agree/use.social.convention	-4.995486878059192
respond.agree/deflection	-12.936647375752456
respond.agree/respond.agree	-3.728381747945511
respond.agree/give.statement	-0.34331611209982277
respond.agree/inner.dialogue	-1000
respond.agree/closed.question	-0.6306002962640211
respond.agree/exclamation	-1000
respond.agree/open.question	-6.541250560041724
respond.agree/quoted.speech	-9.844839064226152
give.statement/respond.deny	-10.264712397163192
give.statement/give.opinion	-3.978354297285268
give.statement/misc	-8.606413430045414
give.statement/recall	-5.991662917038544
give.statement/relax.atmosphere	-5.178448181002791
give.statement/use.social.convention	-5.678963839579919
give.statement/deflection	-15.405724706690267
give.statement/respond.agree	-5.815994687746143
give.statement/give.statement	-0.8353939539911001
give.statement/inner.dialogue	-17.974919891553107
give.statement/closed.question	-6.3045522872802495
give.statement/exclamation	-10.36277256150572
give.statement/open.question	-12.262467655710035
give.statement/quoted.speech	-11.688311909945462
inner.dialogue/respond.deny	-1000
inner.dialogue/give.opinion	0.8880411374896692
inner.dialogue/misc	0.32138254939366584
inner.dialogue/recall	1.73177456389435
inner.dialogue/relax.atmosphere	-1000
inner.dialogue/use.social.convention	-1000
inner.dialogue/deflection	-1000
inner.dialogue/respond.agree	-1000
inner.dialogue/give.statement	1.7023265534558896
inner.dialogue/inner.dialogue	-0.4791678366985595
inner.dialogue/closed.question	-1.0331034417990912
inner.dialogue/exclamation	-2.4032677223393013
inner.dialogue/open.question	-1000
inner.dialogue/quoted.speech	-1000
closed.question/respond.deny	-8.103599548290672
closed.question/give.opinion	-2.3115161775015145
closed.question/misc	-5.7526438835136595
closed.question/recall	-3.6945536129438556

closed.question/relax.atmosphere	-2.673791558684629
closed.question/use.social.convention	-3.518544260802858
closed.question/deflection	-12.157676221046465
closed.question/respond.agree	-3.909221503123821
closed.question/give.statement	0.07013767917112361
closed.question/inner.dialogue	-13.504478984572856
closed.question/closed.question	-2.5545888516776376
closed.question/exclamation	-8.062256655967783
closed.question/open.question	-9.961951750172098
closed.question/quoted.speech	-9.802833503686367
exclamation/respond.deny	-5.557673279148827
exclamation/give.opinion	0.23964674000647765
exclamation/misc	-2.104619426753078
exclamation/recall	-2.5015823343099974
exclamation/relax.atmosphere	-0.8387685871227933
exclamation/use.social.convention	-1.153453767313775
exclamation/deflection	-9.995078591456124
exclamation/respond.agree	-3.084493512289865
exclamation/give.statement	1.7908977501389045
exclamation/inner.dialogue	-9.756918854261361
exclamation/closed.question	-0.7428983839464258
exclamation/exclamation	-1.8552649070189893
exclamation/open.question	-8.214391619860601
exclamation/quoted.speech	-8.225198374817184
open.question/respond.deny	-1000
open.question/give.opinion	1.1249147635755126
open.question/misc	-1.441743824520491
open.question/recall	-0.18335490346485636
open.question/relax.atmosphere	0.02942592308729991
open.question/use.social.convention	-0.3530746413312532
open.question/deflection	-1000
open.question/respond.agree	-0.12883616082943214
open.question/give.statement	1.617272084654371
open.question/inner.dialogue	-1000
open.question/closed.question	-0.6587262919633129
open.question/exclamation	-1000
open.question/open.question	-2.4222330006830477
open.question/quoted.speech	-6.076895945414355
quoted.speech/respond.deny	-5.709685434720195
quoted.speech/give.opinion	0.7314907742098343
quoted.speech/misc	-1.1570959087735315
quoted.speech/recall	-1.3316663949940033
quoted.speech/relax.atmosphere	-2.756315489057138

quoted.speech/use.social.convention	-1.4689646551680224
quoted.speech/deflection	-1000
quoted.speech/respond.agree	-2.1776119788076644
quoted.speech/give.statement	2.6388855945675362
quoted.speech/inner.dialogue	-1000
quoted.speech/closed.question	-2.5115818999662882
quoted.speech/exclamation	-5.466708681227654
quoted.speech/open.question	-1000
quoted.speech/quoted.speech	-1.222392421336448

Table 15: Insertion score, right given left, over all traces.

Traces	Similarity
1 and 2	354.24397550048366
1 and 3	-60.86093826124067
1 and 4	786.5433860682696
1 and 5	-47.53860439947408
1 and 6	305.63823934158944
1 and 7	-84.09637676670282
1 and 8	318.49673690358736
1 and 9	342.45523411196393
1 and 10	-494.7322200808454
1 and 11	-999.6921134194079
1 and 12	-1139.8171195089988
2 and 3	-511.03495867069955
2 and 4	291.33566785825735
2 and 5	-182.58787829201307
2 and 6	163.8639639169548
2 and 7	-96.21268761972284
2 and 8	189.75903536819976
2 and 9	92.29172813022076
2 and 10	-637.1147861500058
2 and 11	-981.3144328364158
2 and 12	-1166.0801019386984
3 and 4	39.255800406738345
3 and 5	-950.0240279635962
3 and 6	-517.0356885450141
3 and 7	-73.20836186893754
3 and 8	-554.3835814226758
3 and 9	-613.5821086783856
3 and 10	-700.6104089075229
3 and 11	-939.9759594930729
3 and 12	-1473.4916473733633

4 and 5	-284.28552223483706
4 and 6	231.3205550543267
4 and 7	-9.650791125720003
4 and 8	319.36316988809176
4 and 9	323.77377922104154
4 and 10	-449.93464512379086
4 and 11	-874.5283831031115
4 and 12	-1060.221531083582
5 and 6	-487.70239770784406
5 and 7	-1122.968491693879
5 and 8	-489.65342270931166
5 and 9	-652.0133562225868
5 and 10	-2277.2304391118582
5 and 11	-1765.015323295438
5 and 12	-3338.3686214204795
6 and 7	25.400184705228735
6 and 8	151.53999144930978
6 and 9	146.00340414428138
6 and 10	-548.2511010922423
6 and 11	-988.4319933078347
6 and 12	-1203.8269414588724
7 and 8	22.267960973722335
7 and 9	28.752296577028417
7 and 10	-335.1818494131894
7 and 11	-909.6342483837044
7 and 12	-643.9323751558169
8 and 9	233.90910188041437
8 and 10	-654.6475547391527
8 and 11	-935.6917930890912
8 and 12	-1074.3155446053186
9 and 10	-583.707451005071
9 and 11	-948.1604190934016
9 and 12	-944.0963202934992
10 and 11	-241.23559446433734
10 and 12	-421.1789257655095
11 and 12	-872.539584486338

Table 16: Similarity between all traces.

Pattern	Rank Score
give.statement, give.statement, give.statement, give.opinion	29.185661610679364
give.opinion, misc	28.476608295358854

exclamation, give.statement	26.254804992236714
closed.question, give.statement	24.5001135248475
misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc	24.469656581723036
misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc, misc	24.469656581723036
recall, misc	23.907674116578782
give.statement, exclamation	23.907674116578782
give.statement, give.statement, closed.question	22.74667515491638
...	...
recall, recall, misc	21.41719617888822
exclamation, exclamation	21.41719617888822
closed.question, closed.question, closed.question	21.41719617888822
closed.question, closed.question	20.22463057322567
misc, misc, misc, misc, misc, misc, misc	19.939963357229512
give.opinion, open.question	18.759271266376626
give.opinion, give.opinion, misc	18.759271266376626
give.statement, give.statement, give.opinion, give.opinion	18.759271266376626
...	...
quoted.speech, quoted.speech	17.07691008327056
...	...
recall, give.opinion, give.opinion	15.899991218358139
give.opinion, give.opinion, give.opinion, give.statement	15.899991218358139
give.statement, give.statement, give.statement, give.statement, give.opinion	15.899991218358139
closed.question, exclamation	15.899991218358139
misc, misc, misc	15.184764653707179
give.statement, closed.question	14.46636000758086
relax.atmosphere, relax.atmosphere, closed.question	14.27813078592548
misc, misc, give.opinion	14.27813078592548
give.statement, misc	13.661528066616446
give.statement, give.opinion	13.059583407257833
open.question, respond.agree	12.788039631123947
respond.agree, recall	12.788039631123947
give.opinion, exclamation	12.788039631123947
give.opinion, recall, recall	12.788039631123947
recall, recall, recall, give.opinion	12.788039631123947
give.opinion, give.opinion, give.opinion, use.social.convention	12.788039631123947
give.opinion, give.opinion, give.opinion, misc	12.788039631123947
recall, recall, recall, closed.question	12.788039631123947

give.opinion, quoted.speech	12.788039631123947

Table 17: Sequential Patterns from PrefixSpan variation.

References

[1] DENIS KUPERBERG. "LINEAR TEMPORAL LOGIC FOR REGULAR COST FUNCTIONS". In: *LOGICAL METHODS IN COMPUTER SCIENCE* (Feb. 8, 2017). DOI: 10.2168/LMCS-???.

[2] Guillaume Lample et al. "UNSUPERVISED MACHINE TRANSLATION USING MONOLINGUAL CORPORA ONLY". In: *ICLR 2018* (Apr. 13, 2018). DOI: arXiv:1711.00043v2.

[3] Ian J. Goodfellow et al. "Generative Adversarial Nets". In: (June 10, 2014). DOI: arXiv:1406.2661v1.

[4] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. "Sequence to Sequence Learning with Neural Networks". In: (Dec. 14, 2014). DOI: arXiv:1409.3215v3.

[5] Jian Pei et al. "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth". In: *Proceedings 17th International Conference on Data Engineering* (Apr. 2, 2001). DOI: 10.1109/ICDE.2001.914830.

[6] Majid Rafiei and Wil M.P. van der. "Towards Quantifying Privacy in Process Mining". In: (Dec. 21, 2020). DOI: arXiv:2012.12031v1.

[7] Majid Rafiei and Wil M.P. van der Aalst. "Privacy-Preserving Data Publishing in Process Mining". In: (Jan. 4, 2021). DOI: arXiv:2101.02627v1.

[8] Marsha Chechik and Dimitrie O. P˘aun. "Events in Property Patterns". In: (June 29, 1999). DOI: arXiv:cs/9906029v2.

[9] Nijat Mehdiyev and Peter Fettke. "Explainable Artificial Intelligence for Process Mining: A General Overview and Application of a Novel Local Explanation Approach for Predictive Process Monitoring". In: *Springer* (Oct. 7, 2020).

[10] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. "Natural language processing: an introduction". In: *J Am Med Inform Assoc* (2011). DOI: 10.1136/amiajnl-2011-000464.

[11] R. P. Jagadeesh Chandra Bose and Wil M.P. van der Aalst. "Context Aware Trace Clustering: Towards Improving Process Mining Results". In: *SIAM International Conference on Data Mining* (Apr. 2009). DOI: 10.1137/1.9781611972795.35.

[12] Tom Young et al. "Recent Trends in Deep Learning Based Natural Language Processing". In: (Nov. 25, 2018). DOI: arXiv:1708.02709v8.

[13] Tomas Mikolov et al. "Efficient Estimation of Word Representations in Vector Space". In: (Sept. 7, 2013). DOI: arXiv:1301.3781v3.

[14] Wil Knight. "AI's Language Problem". In: (Aug. 9, 2016). URL: https://www.technologyreview.com/2016/08/09/158125/ais-language-problem/.

[15] Wil van der Aalst. *Process Mining Data Science in Action.* 2nd ed. Springer, 2016. 467 pp. ISBN: 978-3-662-49850-7.