Exploring Role of Interjections in Human Dialogs

Jean-Baptiste de la Broise
EPFL
jean-baptiste.delabroise@epfl.ch

Ekaterina Svikhnushina
HCI-EPFL
ekaterina.svikhnushina@epfl.ch

Abstract

Interjections are words and expressions that people use to communicate sudden reactions, feelings, and emotions. Most of the time, the usage of interjections is unconscious, and they are a primary part of a human conversation at the border between verbal and non-verbal communication. Naturalness is a primary aspect for open-domain conversational agents. Therefore the role of interjections for chatbots' perceived naturalness is under investigation. Earlier studies demonstrated that chatbots are evaluated as more natural and engaging when they use interjections in their responses. In this context, we seek to develop a large and convenient taxonomy of interjections, which the chatbots could easily use. We will evaluate the taxonomy using two data sets: the goEmotions data set composed of Reddit comments annotated with emotions and the empathetic dialogues data set containing conversations bound in an emotional context. We will derive a data set from the second one to allow the creation of a model able to predict which interjection to add to improve a chatbot answer to a given utterance.

1. Introduction

1.1. Description of the challenge

Since the Turing test invention, Mankind tries to find ways to create an agent able to pass this test, an artificial agent able to communicate as naturally as a human would. We want here to bring a little step to this quest. We recently got evidence that the usage of interjections improves chatbots' perceived naturalness [2], but interjections is a very complex part of a language. We can witness this fact, as interjections are probably the hardest part of a language for a learner to grasp the proper use [7]. It is hard to explain the proper use of an interjection as this feels so natural to the native speaker, yet it is often very context-dependent. This is what makes interjections an interesting learning task for a language model. We make the following hypotheses:

- We can derive a large taxonomy of interjections.
- We can split these taxonomy interjections into classes to alleviate some of their complexity.
- We can grasp some of the taxonomy interjections' use complexity with proper visualizations.
- We can derive a data set using some existing data-sets to simplify the training for a classification task involving proper interjections' use.

1.2. Purpose of this work

We will create a taxonomy of interjections. We try to select those interjections for their propensity to make an impact on chatbots' perceived naturalness. To understand interjections complexity better, we first want to describe interjections' use in different contexts. We also want to discover how to bring an agent to grasp the interjections' use better. We want to find out if there is a way to design a data set to train a model to predict which interjection to use in response to a given context.

1.3. Steps to follow

1. Create a near-complete taxonomy of interjections.
2. locate those interjections in the data sets we work with [3] and [6]. Visualizing the obtained data in a way that allows us to get insights on interjections' proper use. (position in sentences, the emotional context associated with each, sentiments associated with each)
3. Reduce the taxonomy by keeping the most relevant interjections which are widely present in the data sets.
4. Design a data set for a classification task to identify which interjection fits to respond to a given utterance.

2. Taxonomy of interjections based on literature review

We started by designing a broad taxonomy of interjections[1]. To build this taxonomy, we use the interjections

[1] The taxonomy is made available at this link: `https://docs.google.com/spreadsheets/d/1IDXYgBFH2_RIyyF0iS6qYuMdj3eRcQaARINch3yd1yo/edit?usp=sharing`

we encountered in the papers we studied. We added some found in a range of online dictionaries. We started by adding an informal definition and a use case description for each of the interjections. It will help us grasp the meaning of the interjections we study, which is very important as some of us are non-native English speakers. Those definitions will prove to be useful in the paragraph 3.4.3.1 as that work will give us the ability to group interjections with similar meaning. We want to be able to split interjections into some subclasses in order to characterize those better. We, therefore, classified the interjections using different categorisations encountered in the papers.

2.1. Primary and secondary interjections

Found in the book [5] this classification splits the taxonomy into two categories. The primary interjections category members do not have any other role than an interjective role. This classification is easy to make as we can check in dictionaries to see if the interjection has other linguistic use. This category includes the following interjections: "Ahem", "ouch", "wow!" Secondary interjections, as opposed to primary ones, have other uses than the interjective one. Secondary interjections are hard to study, as, to do it properly, you should first find a way to discriminate between the roles those can arbor depending on the context. Secondary interjections contains:

- "Awesome" which is also an adjective, "well" which is also an adverb
- "Well" which is also an adverb

2.2. Expressive, Conative, Phatic, and Routines

This classification is taken from the two following sources [4] and [5].

The expressive class contains two sub-classes, the emotive *e.g.* "ouch" and cognitive *e.g.* "oh". We decide not to use those sub-classes in our classification as we feel those do not bring much more insight than the "expressive" upper class. This class is an interesting one to study. We can assume it contains most of the impactful interjections for perceived naturalness.

- The conative class members target an addressee and call for a behavior. *e.g.* "Shh!" and "pst!"
- The phatic class interjections use is the maintenance of a communicative contact. While listening to show you are attentive or to get the speaker's attention to take the conversation lead. Or while speaking when you are hesitating. *e.g.* "er" and "hm" to use as the speaker, "Nah" to show disapproval as the listener. The phatic class seems hard to teach a chatbot the proper use, as many of those appear at random timing throughout a conversation.
- Routines: *e.g.* "What's up!" and "bye-bye"

2.3. Valence and Emotional orientation

This classification composed of two sub-classification we found in [2] The valence classification is composed of three classes:

- The positive class: This class is composed of interjections such as *e.g.* "Fantastic", which we deem as positive.
- The negative class: This class is composed of interjections such as *e.g.* "D'Oh!" which people use in a negative context for you or another person *c.f.* emotional orientation
- The neutral class: This class contains interjections that have a positive or negative use, "no way!" is one of those. It also contains interjections like "pst!" that are neutral as they are neither used positively nor negatively.

The Emotional orientation classification contains two classes:

- self-oriented: Interjections that people use when speaking alone or speaking to somebody while being the main element of the conversation. "Aha!", "D'Oh", "Oh"
- other-oriented: This category contains Interjections used to show your understanding of your interlocutor sentiments. you can also use those to place your interlocutor on a pedestal showing his importance in the conversation. *e.g.* "Oh dear", "wow!", and "you bet!"

Emotional orientation is one of the hardest classifications we encountered. We believe the Emotional orientation classification is quite fuzzy. This classification is interesting as it was used within the Alexa prize and give some results we can use in this work. For example, Interjections oriented towards others have a positive effect on perceived naturalness. We should try to study those other-oriented interjections further. Valence unexpectedly made no difference in perceived naturalness during this study.

2.4. Classification process

For each classification method above, To classify an interjection we:

1. Look at the paper presenting the classification to see if it is already classified. If it is, we follow the authors' class choice. Search for an interjection classified in the paper and which has a similar meaning or use. If we find a match, we give the same class chosen for the paper's interjection to the interjection we seek to classify.

2. Look at the class definitions and search for a match between one of the class definition and the interjection definition.

3. In some cases, the class choice was not obvious enough for us to make any assumption, in which case we left the field blank.

This classification will allow us to select reliable classes to keep in the final taxonomy to create the data set for the classification task. Indeed some classes are hardly usable. As an example, we can talk about secondary interjections for which it is hard to distinguish between the interjection role and other roles. We tried to use NLTK Part of Speech tagging (PoS) to do this, but it seems not very reliable for interjections tagging.

interjection	use or definition	2.1	2.2	Valence 2.3	Orientation 2.3
Sh	requesting silence	primary	conative	negative	other
ouch	express pain	primary	expressive	negative	other
Oops	acknowledging a mistake	primary	expressive	negative	self

Table 1: Sample of the taxonomy. Green cells are the ones for which we found the class in the related paper. Blue cells are the ones for which we are pretty confident in our classification.

3. Data Analysis

3.1. Interjections location

We want to get the ability to locate the interjections found in our taxonomy in pieces of text to make subsets out of other data sets. for each interjection in the taxonomy, We create a regular expression that can match this interjection and its variations. *e.g*. regular expression for interjection 'ah' matches 'ah', 'ahhh', ... The regular expression for interjection 'oh' matches 'oh' but does not match 'oh boy' which is another interjection found in the taxonomy.

3.2. Data Sets

3.2.1 goEmotions

The goEmotions [3] data set contains annotated comments scraped from Reddit. We chose this set as we think we can find many interjections here because social networks tend to lead to very emotional posts. This data set will allow us to discover which emotions can be associated with each interjection. It will also allow us to witness the differences and similarities between the conversational and the blog-post formats. Interjections are often very expressive, which makes them well suited for this context. We think this effect could be prominent on Twitter, where you try to express the most with a limited amount of characters. (Interjections seems to be very relevant in that context). Annotators gave each comment an emotional label. The label is within an emotional range (27 emotions and neutral). We decide to make a subset out of the goEmotions set. We do so by locating each post containing interjections (122 interjections were first selected). The original goEmotions data set contains one row per annotation, leading to multiple rows for the same comment. We decide to group the rows by text content. We then make a poll to aggregate the annotations.

3.2.2 Empathetic dialogues

The empathetic dialogues data set [6] is a data set referencing 25'000 conversations with the emotional context associated. The conversations are composed of utterances. A conversation is an alternation between a 'speaker' and a 'listener' utterances. The conversation begins with a 'speaker' utterance. We use this data set in addition to the goEmotions set. As it is referencing conversations, it will allow us to:

- Analyze the data from the speaker and listener point of view. (chatbots being very often in the listener role, this split is interesting).
- Create a data set to train a model to predict which interjection to use to react to an utterance.

We modified the data set slightly by grouping some emotions and tweaking some punctuation.

3.2.2.1 Emotions grouping

We grouped the emotional contexts in classes (we call those classes upper-contexts). We thought that contexts were originally not selected in a meaningful way (not based on any classification), and the number of contexts was too high for proper visualization. We base our emotion upper-contexts on Plutchik's wheel of emotion Figure 1. Originally there were 32 emotional contexts. We select our upper-contexts from the middle ring of the emotion wheel (Anger, Disgust, Anticipation,...). This ring contains emotions with not too much, nor too little intensity, which allows an easier categorization of those (*e.g*. Terror we feel is too strong an emotion to describe an upper context. Annoyance, on the other hand, is an emotion which is too light to define an upper context). To this selection, we added some of the mixed-emotions (Remorse, submission,...). We end up with 14 upper-contexts we could map to at least one of the 32 original emotional contexts *e.g*. the upper-context sadness gathers the following original contexts: devastated, disappointed, lonely, sad, sentimental, and nostalgic. The upper-context 'remorse' contains only the 'guilty' context.

To map a context to an upper-context, we:

- try to find an emotion similar to the context in the Plutchik's wheel.

- If we find a match, we affect the context to the closest upper-context to the found emotion.

- Else, we have to decide the mapping by ourselves. We try to figure out the best match based on the background bits of knowledge we have on the emotion topic.

We are aware our new distribution is questionable as it presents an imbalance between the upper-contexts. We think it is not a problem as we do not plan on training a model with emotional context as a label. We shall be careful and remember we cannot analyze the number of interjections used in each context without normalization. Our mapping is open to discussion, as some people might want some of the original contexts mapped to other upper contexts.

3.2.2.2 punctuation formating

We removed some punctuation, such as the '_comma_' which we felt would not be usable by an embedding system. Our subset contains 52'691 utterances found in 23'152 conversations.

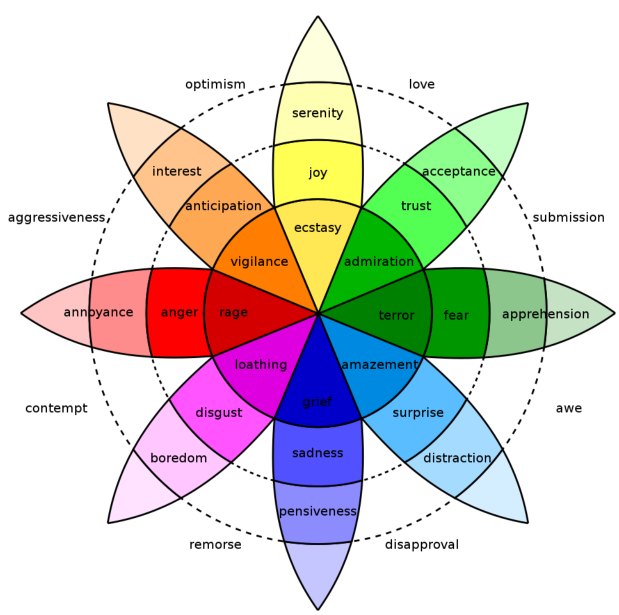


Figure 1: Plutchik's wheel of emotions Wikimedia Commons under the license of CC Public Domain Mark 1.0

We then did the same process as for the goEmotion data set to find all of the interjections present in the utterences.

3.2.3 Resulting sets

We obtain two subsets that are very imbalanced, some interjections being way more frequent than others. For example, as we use regular expressions, secondary interjections interjective role is not discriminated from their other roles. It leads to many secondary interjections like 'really' being way more represented than primary ones, as we can see in the table 2. A better way to get a balanced data set could be to build a data set from scratch, scraping data from the web. By doing so, we could force a balance. It could also allow us to keep only interjections identified as such by a PoS tagging model, as the amount of available data would not be a problem. It would require way more data than we have convenient access to. It will also not (or hardly) allow us to get a conversational set like the empathetic dialogues set.

Empathetic dialogues

ta	1
horay	1
high five	1
uh-oh	1
poh	1
...	
oh	2861
great	3525
no	4581
really	5109
so	10208

goEmotions

what a man	1
sho	1
aw yeah	1
ta	1
as you wish	1
...	
well	1247
lol	1513
really	1895
no	2473
so	4489

Table 2: frequency of several interjections in both data sets

3.3. Methods

For both sets, an in depth analysis was conducted to better identify how the interjections are used in social networks comments and conversations. Those analysis include:

- analysis based on the labels given by the data sets (Emotions for goEmotion, conversations emotional context and speaker/listener split for empathetic dialogues)

- interjections relative positioning in the texts

- interjections collocations and most frequent n-grams

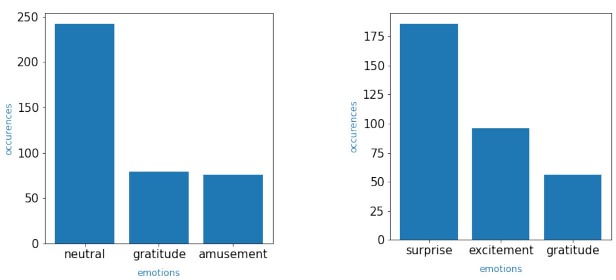
3.4. Results

Using the method we defined in the previous subsection, we derived some visualizations to get some insights.

3.4.1 Use of interjections in Reddit comments

3.4.1.1 Interjections related to emotions

For each interjection, we made a plot showing the top 3 emotions associated with it. Neutral being an emotion allows us to witness a separation between heavily emotionally involved interjections and less involved ones. We analyzed primary and secondary interjections apart. We do so because, for the reasons stated in subsection 3.2.3, for secondary interjections, the analysis is not as reliable.

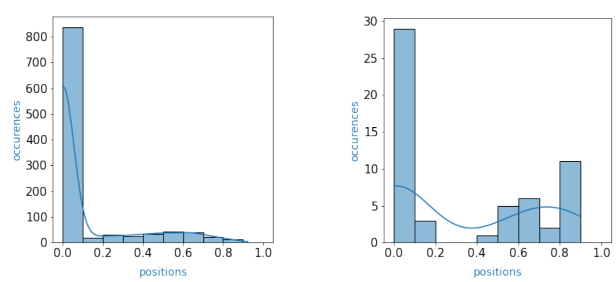


(a) 'oh' (b) 'wow'

Figure 2: Emotions associated with two primary interjections, (a) showing a neutral interjection, (b) showing an emotionally involved interjection

3.4.1.2 Position of interjections

For each interjection, we made a plot showing the interjection position distribution in sentences.



(a) 'oh' (b) 'yikes'

Figure 3: relative Positions of two primary interjections in a, (a) one-sided distribution, (b) more balanced distribution

We can see that both interjections shown in the Figure 3 are often in the first 10% of a comment. The kernel density estimation allows us to be more precise and to say that most of the time, those interjections are in position 0, which corresponds to the first word of the comment. We witness this trend among many interjections. Those results Figure 3 are interesting to use in a learning task where the goal is to position a given interjection in a given sentence. A decision tree would probably always decide to position interjection 'oh' at the beginning of a sentence, regardless of the given sentence. Note that results may vary in the conversational context, we will witness those variations or similarities in 3.4.2.2.

3.4.1.3 n-grams associated with interjections

The study of n-grams associated with an interjection can lead to some insights. It allows us to get an idea of which words should be in a sentence containing said interjection.

lol	wow	damn
('in', 'the'), 57	('this', 'is'), 27	('i', 'love'), 17
('i', 'was'), 47	('what', 'a'), 18	('i', 'was'), 15
('of', 'the'), 37	('for', 'the'), 15	('name', 'i'), 13
('it', 'was'), 34	('in', 'the'), 15	('in', 'the'), 12
('on', 'the'), 32	('thank', 'you'), 14	('that', 'was'), 12
('i', 'think'), 28	('so', 'much'), 13	('this', 'is'), 11
('for', 'the'), 26	('oh', 'i'), 13	('it', 'i'), 10
('i', 'love'), 25	('that', 'was'), 12	('name', 'it'), 10
('thank', 'you'), 25	('i', 'am'), 12	('what', 'a'), 9
('this', 'is'), 25	('thanks', 'for'), 12	('for', 'a'), 9

(a) no filter

lol	wow	damn
('dont', 'know'), 8	('thank', 'much'), 6	('pretty', 'good'), 6
('feel', 'like'), 8	('dont', 'know'), 6	('looks', 'like'), 5
('im', 'sure'), 8	('never', 'heard'), 5	('every', 'time'), 3
('sounds', 'like'), 7	('didnt', 'know'), 5	('could', 'use'), 3
('im', 'glad'), 6	('sounds', 'like'), 5	('didnt', 'know'), 3
('didnt', 'know'), 6	('thats', 'amazing'), 4	('sorry', 'hear'), 3
('every', 'time'), 6	('oh', 'didn't'), 4	('thats', 'good'), 3
('looks', 'like'), 6	('didn't', 'realize'), 4	('i'd', 'love'), 3
('seems', 'like'), 5	('never', 'seen'), 4	('feel', 'lol'), 2
('made', 'laugh'), 5	('oh', 'thats'), 4	('brazen', 'enough'), 2

(b) stop words filtered

Table 3: top 10 2-grams for some interjection. in the second table we filtered NLTK english stopwords and the word 'name' which was over-represented because of the data anonymization process.

pretty	people	really
85	39	32

Table 4: top words of size ≥ 6 occurring with secondary interjection 'sure'

Table 4 emphasize the problem with secondary interjections. indeed we can see the word "pretty" being often collocated with "sure" probably to form "pretty sure" which is not an interjection. This shows that the primary use of 'sure' is not its interjective use. Consequently, The way "sure" and many secondary interjections are studied here leads to results that are not about interjections but words. It might be a problem for the construction of a training data set subsection 4.1.

3.4.2 Use of interjections in empathetic dialogues

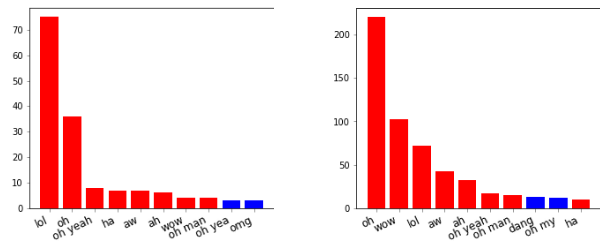
On the empathetic dialogues data set, We conduct an analysis similar to the one we did on goEmotions. The main differences are :

- The analysis tackles the differences between the conversation speaker and listener.
- The emotional contexts are different than the emotion from goEmotions.

- removal of some interjections to focus on 72 expressive interjections

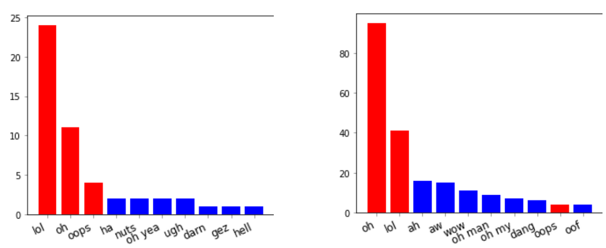
3.4.2.1 Relationship between emotional context and interjections' use

For each emotional context we make a figure representing the main interjections use by conversation speaker and listener in that context.



(a) speakers (b) listeners

Figure 4: Interjections use in the "joy" emotional context. A red bar signifies the interjection is part of the top 10 on both sides.



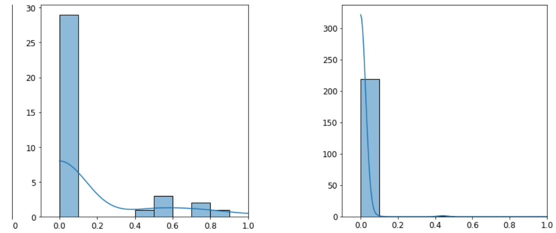
(a) speakers (b) listeners

Figure 5: Interjections use in the "remorse" emotional context.

This observation allows us to notice some contexts where interjections used by speakers and listeners are very similar *e.g.* in the context "joy" (Figure 4). In some other contexts like the "remorse" context, interjections' use by speaker and listener varies significantly Figure 6. Those results might be due to some context calling for a share of sentiments between speaker and listener wherein some other contexts, the emotions the listener and speaker experiments should differ. Joy is easy to share, where remorse is more self-related.

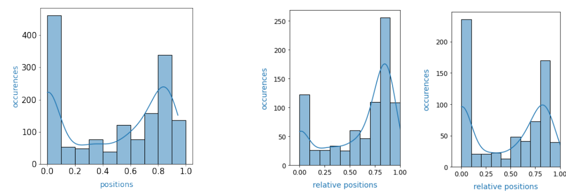
3.4.2.2 Position of widely used interjections per context

For each emotional context, we study the relative position of the most used interjections in the context.



(a) Speakers (b) Listeners

Figure 6: Position of interjections "oh" used in the "joy" emotional context.



(a) 'Reddit' (b) 'Speaker' (c) 'Listener'

Figure 7: relative Positions of primary interjections 'oh' in: (a) Reddit comments, (b) empathetic dialogues speaker utterances (c) empathetic dialogues listener utterances

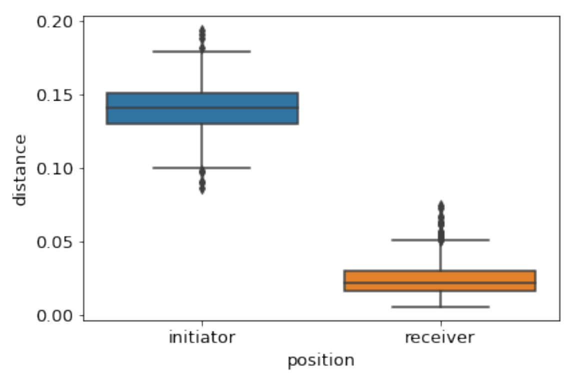


Figure 8: Wasserstein distance 'listener' distribution and 'speaker' distribution with Go distribution confidence interval for interjection 'lol' with Bootstrap confidence intervals 95%

For some interjections, we notice differences between where the speakers and listeners put those. We believe those differences are significant. Most of the listeners' interjections placement is at the beginning of the utterance. This interjections' positioning is probably due to the listener assessing his understanding of the speaker's feelings and validating his emotions. Interjections are certainly powerful

emotional speech elements. Proper use of interjections is allowing the creation of an empathetic relationship between speakers and listeners.
The Figure 7 let us witness some interesting correlations between the distribution of some interjections such as 'lol'. Indeed, interjection 'lol' seems to be pretty much used the same way in the Reddit comments context than in the conversational 'listener' context. We would need further study to see if we can witness such correlations for other interjections. We can guess the Reddit commentator position is a reactive position, which might have similarities with the conversational listener position. Figure 8 allows us to say that the distribution of interjection 'lol' in Reddit comments and in the conversational listener context are closely correlated.

3.4.2.3 n-grams associated to context and interjections

We studied one and two grams. We can get some insights into the language and roles of listeners and speakers in the conversation. For many contexts, where the listener can empathize, the word 'you' is very present as a call for a share of emotions. The listeners, in the joy emotional context, tend to use interjections categorized with emotional orientation:'other'. *e.g.* The interjection 'wow' is widely used by the listeners in the 'joy' emotional context. Listeners use 'wow' in collocation with 2-grams like 'are you', 'did you', and 'that is'. Those 2-grams use is, in general, to put the speaker on a pedestal to let him further express his sentiments. In some other context such as 'remorse', listeners tend to use 'that' more than 'you', this use might be a way to show an understanding of the situation and a way to try not to blame the speaker lightening his burden in the process. The vocabulary in this context includes 2-grams like 'to support', 'feel bad', 'it happens'. This vocabulary is used with some interjections like 'oh my' or 'oh man', 'dang' to show compassion.

3.4.3 Further study in the listener context

A chatbot is often put in a listener position. This is why, We want to study a bit more the data we have in the listener context. This study is directly preceding the training data set creation subsection 4.1. We want to have a clear list of interjections to use in our data set.

3.4.3.1 Frequency of interjections over the data set

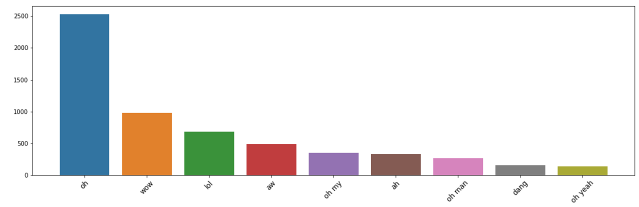


Figure 9: Frequency of interjections with more than 100 occurences in listener context

As we can see in Figure 2, we have a very imbalanced data set, and there is not much we can do at this point to change that. If we had much more data we could have taken samples, here we only have nine interjections with more than a hundred occurrences, so it is not very doable. We will later try some resampling anyway, but we have to keep in mind the imbalance and amount of data here makes this data not very usable. We try to compensate a bit by grouping some interjections with similar meaning Figure 10. this grouping is made possible because we recorded the approximate meaning of each interjection we added to the original taxonomy. Even with the groups, we do not achieve to get much more data, and we fear that by grouping further, we might lose the relevance of the interjection classes.

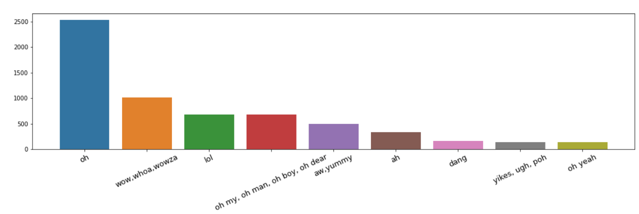


Figure 10: Frequency of interjection-groups with more than 100 occurences in listener context

We also did a Sankey diagram to visualize the groups of interjection distribution over emotional contexts. Figure 11 is not very readable without the interactivity allowed by the package Bokeh. We can all the same witness that the imbalance is not only in the interjections' use but also in the emotional contexts. We should consider that if we want to have emotional context as a feature for the classification task. This diagram also allows us to witness some interjections being very context specific. *e.g.*'ew' is only used in context 'disgust' and 'ah' is only used in context 'sadness'.

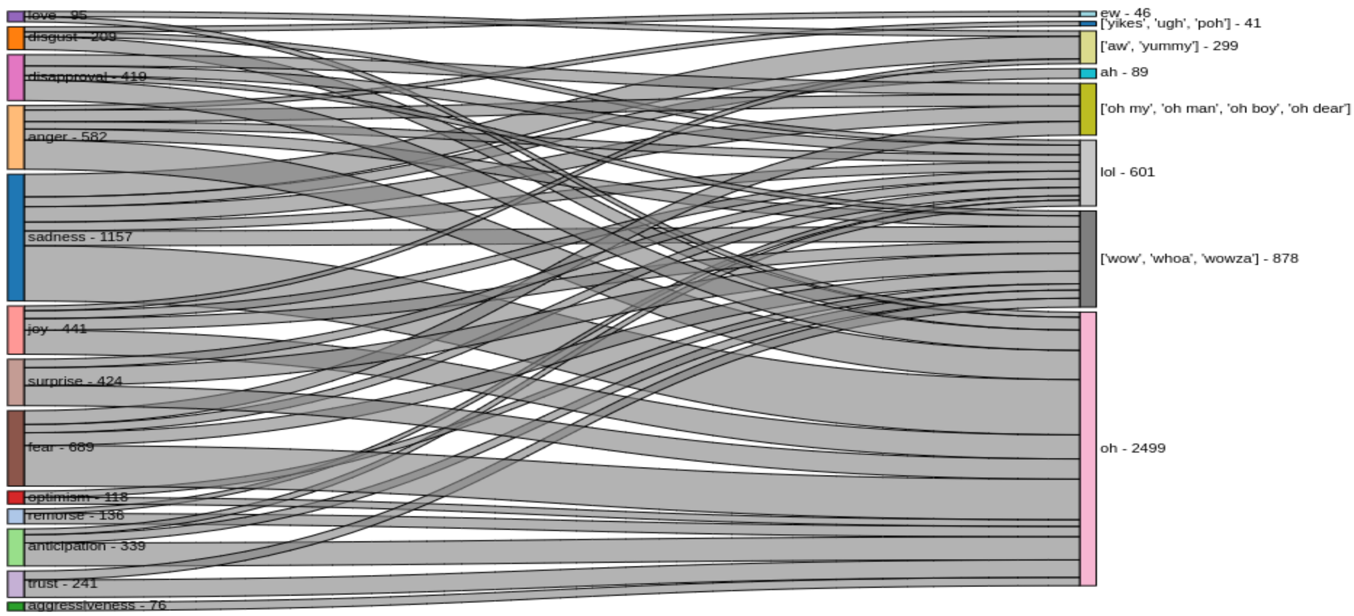


Figure 11: Sankey diagram interjections' use in different contexts

4. Training task: which interjection to use reacting to an utterance

Ultimately the task we try to achieve is the following, How can we add relevant interjections in a given utterance's answer? We can split this task into three subtasks.

1. Should we add an interjection to the given utterance's answer? (single-class classification)
2. If yes, Which interjection is the most appropriate? (multi-class classification)
3. Where in the answer should the interjection be put? (regression?)

We decide to focus on the second task.

We can note that all of those tasks are tackled at once by a language model. Our work is not made useless by the recent breakthrough in language models. Indeed, those models' training requires massive amounts of data to give impressive results. The training process can be very costly *e.g*. the gpt-3 training cost is estimated to be around 4.6 million USD.

4.1. Creation of the data set

We design the data set to create a model for the second subtask. Looking back, we believe we could have made a data set allowing work on the three subtasks at once. Knowing there is an interjection in a given utterance's answer, we want to predict this interjection. The data set contains two features:

- The speaker utterance preceding the interjection's use.
- The emotional context of the converation

The data set is labeled according to the interjection groups found in paragraph 3.4.3.1.

We split the resulting data set into a training set, a testing set, and a validation set. The data set is very imbalanced. For the training, we will try to use some resampling methods. With the remaining time, We wanted to make a first model not necessary to get outstanding results at the task but to have a first insight on the training process and the adjustments to make knowing the imbalance.

5. Limitations and future work

Late in the project timeline, we decided to test the data set we derived from empathetic dialogues. We tried to build a model for the task we decided to work on subsection 4.1, we did some trials and encountered many difficulties.

5.1. Machine learning approaches

We tried to use many different Machine Learning approaches to this problem. We knew we had to do something to solve the imbalance in the data.

5.1.1 Balancing the data

The data set is not balanced. We want to work with a balanced sample (not necessary. We will also try weighted learning). To pick a sample point, we first select a class at random. We then choose a row belonging to said class at random. We repeat the process n-times to form a sample of size n. This sampling process leads to some rows having multiple occurrences in the sample set and some other rows not appearing anywhere in the sample set. We think it is not the best method we could have chosen because we want to use k-fold cross-validation for the hyper-parameters choice, which means, with this sampling method, we will have duplicates in different folds. It might lead to some over-fitting in the k-fold cross-validation for the grid search. In this case, it is advised to do the hyper-parameter search on the original imbalanced data set before resampling and then to train on the resampled set with the found hyper-parameters. We decided to do so. In Future works, we might want to generate new data to solve this. Generative Adversarial Networks could maybe let us create new unique data points respecting the general data distribution.

For this resampling method, we kept only classes with more than a hundred data points. We chose to make a sample of size $3/5^{th}$ of original set size it will lead to approximately 339 data points per class. As a result, we get a data set composed of 3387 data points, but only 2083 of those points are unique.

5.1.2 Utterance's embeddings

We decided to use a simple BERT pre-trained model to get the BERT classification tokens' embeddings. We chose not to take the fine-tuning path as we did not seek to get outstanding results. We first want to get a predictive model running to know how our sampling method affects the results. We did not use the attention mask, as BERT is optimized to work on sentence pairs. The mask is set to a list of 1 with the same size as the message.

5.1.3 Model

We tried to use various model including, but not limited to:

- scikit-learn neural network model on the sample (MLPClassifier)
- KNN on the sample (in a desperate attend as the number of dimensions is way too high)
- Logistic regression on the imbalanced set using parameters class='multinomial', class_weight='balanced'

None of those models gave usable results, as for each method we gave a try the models were over-fitting and predicting the most frequent label regardless of the input. We did not have time left to deep-dive into figuring that out. Figuring out this problem could be a good thing to do in a future project. We could achieve many things during this project, but there are some points we can improve on and some ideas we have for future projects.

- Find a more interactive, meaningful, powerful way to represent our data to increase the viewer insights on the interjections' use in an easy to grasp manner.
- Create a very accurate model for the second subtask 4
- Modify the training set to allow a train over the three subtasks 4.
- Create a balanced data set. It would probably require to find a massive anonymized conversational text source containing many interjections.
- Organize a competition using this data set.
- Compare the resulting specialized model to complete language models like gpt models (maybe comparing to model trained using negative sampling is a good idea as negative sampling restricted to interjections with utterance size frames seems to be quite close to what we try to do). [1]

6. Conclusion

This work allows us to confirm or deny some of the Hypotheses we did in the introduction 1.

- We could grasp some of the interjections' use complexity. We could get some insight on the topics: interjections related to emotions. interjections positioning in sentences. There is also quite a lot we could not grasp properly, and some analysis we could improve.
- We could not prove that a model can grasp interjections' proper use with our data.

This work is an interesting basis for a reader willing to find out what they can expect trying to teach the interjections' use to a chatbot.

References

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language models are few-shot learners, 2020.

[2] M. Cohn, C.-Y. Chen, and Z. Yu. A large-scale user study of an Alexa Prize chatbot: Effect of TTS dynamism on perceived quality of social dialog. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 293–306, Stockholm, Sweden, Sept. 2019. Association for Computational Linguistics.

[3] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online, July 2020. Association for Computational Linguistics.

[4] C. Goddard. Interjections and emotion (with special reference to "surprise" and "disgust"). *Emotion Review*, 6(1):53–63, 2014.

[5] N. R. Norrick. Corpus pragmatics : a handbook. pages 249–275, 2015.

[6] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy, July 2019. Association for Computational Linguistics.

[7] E. Ting. A dictionary of unorthodox oral expressions for english learners and teachers. 2013.