Describing affect and episteme

A mixed methodology for the study of expressive interjections

Nicolás Rivera - Claudio Palominos

Universitat Pompeu Fabra

November 23, 2023

SCAN ME

GLIF
Formal Linguistics
Research Group

Universitat
Pompeu Fabra
Barcelona

GraC
Grammar & Cognition
Research Group

Outline

Object of Study: Expressive Interjections in Chilean Spanish

- Expressive interjections are interactional particles that express emotive and epistemic content. They serve to communicate the speaker's attitude towards a proposition or event.

Object of Study: Expressive Interjections in Chilean Spanish

- Expressive interjections are interactional particles that express emotive and epistemic content. They serve to communicate the speaker's attitude towards a proposition or event.

- Interjections have been described to be loosely integrated into the propositional structure of the utterance (Wilkins, 1992).

Object of Study: Expressive Interjections in Chilean Spanish

- Expressive interjections are interactional particles that express emotive and epistemic content. They serve to communicate the speaker's attitude towards a proposition or event.
- Interjections have been described to be loosely integrated into the propositional structure of the utterance (Wilkins, 1992).
  - But are they actually?

Object of Study: Expressive Interjections in Chilean Spanish

- Expressive interjections are interactional particles that express emotive and epistemic content. They serve to communicate the speaker's attitude towards a proposition or event.
- Interjections have been described to be loosely integrated into the propositional structure of the utterance (Wilkins, 1992).
  - But are they actually?
- Despite their importance in interaction, little is known about these kind of particles (Ponsonnet, 2022; Dingemanse, 2023).

Object of Study: Expressive Interjections in Chilean Spanish

- There's a need for a methodology to describe expressive interjections.

Object of Study: Expressive Interjections in Chilean Spanish

- There's a need for a methodology to describe expressive interjections.
- Interjections are highly attached to the language's culture (Wierzbicka, 1992). An investigation on interjections must be, at first, culture-specific.

Motivation

- *pucha* and *chuta* are phonetically similar words that speakers use in very similar contexts to express similar content.

Motivation

- *pucha* and *chuta* are phonetically similar words that speakers use in very similar contexts to express similar content.
- Native speakers intuitively recognize that they are somehow different, though they don't know exactly why (Rivera, 2022).

pucha

chuta

Motivation

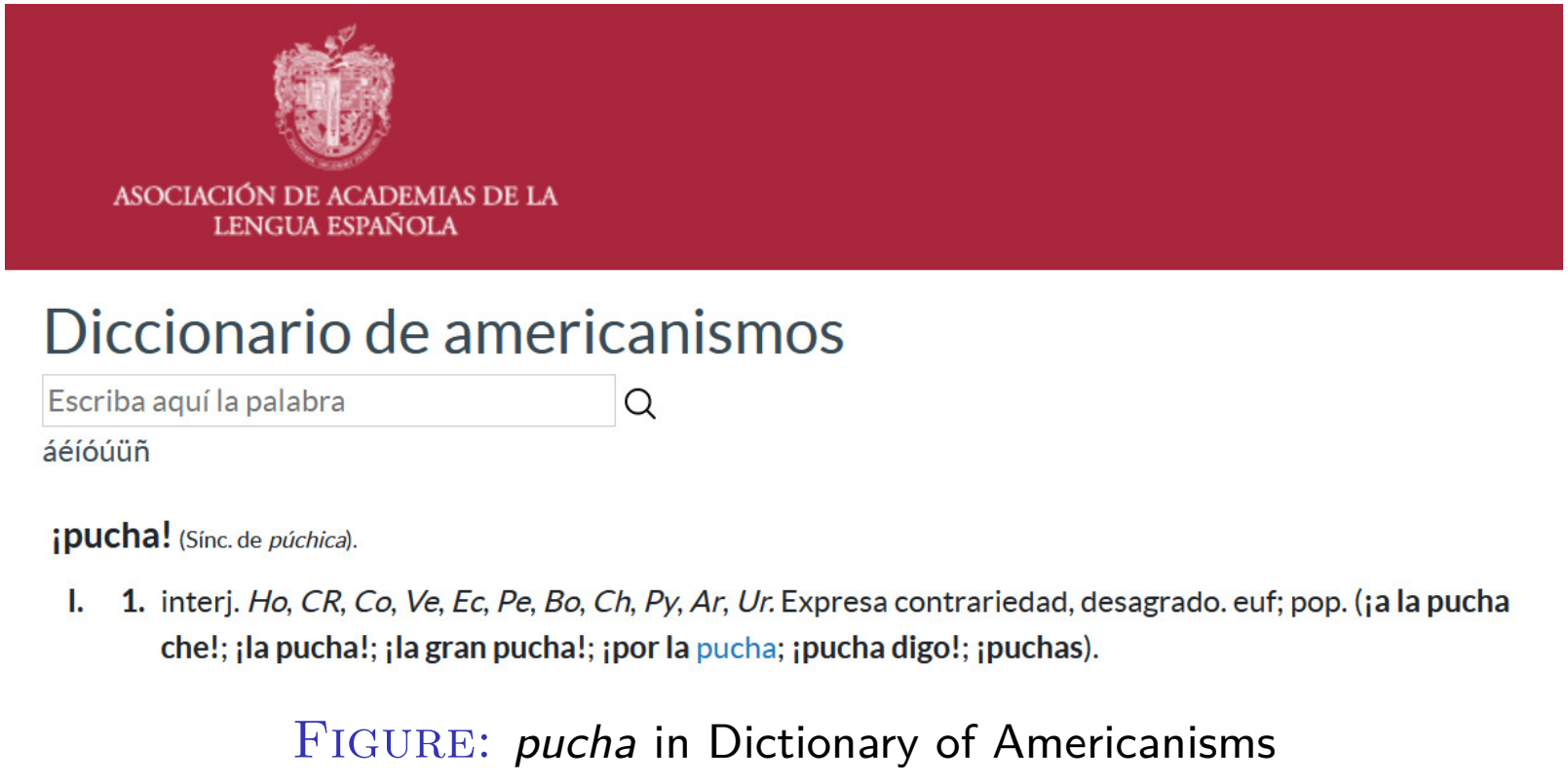


FIGURE: *pucha* in Dictionary of Americanisms

Motivation


Viejo Perro Blanco @ander_zavala · 21h
Ya en casa y como estropajo con el dolor
7
9
304
VanBass
@VBassaletti
Pucha, no sé qué tienes. Espero que sanes pronto y ya no sientas dolor.

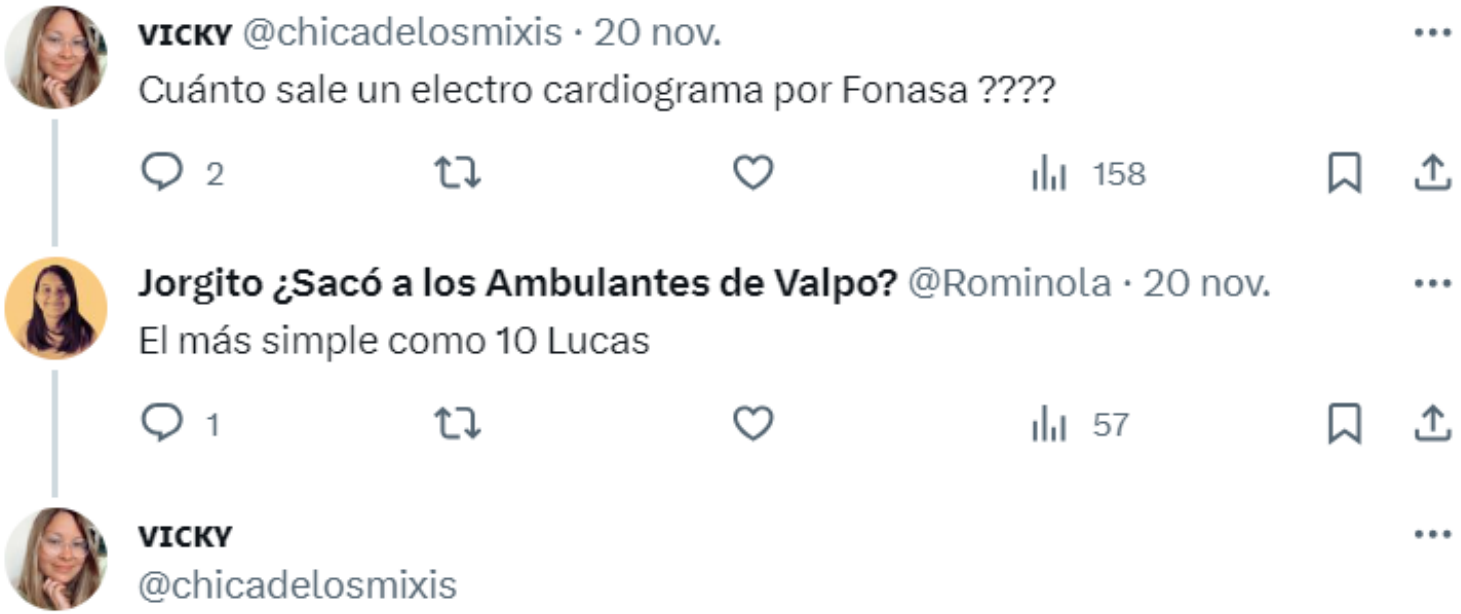
Motivation

I: Already at home and like a scouring pad because of the pain.
R: **Pucha**, I don't know what you have. I hope you heal soon and no longer feel pain.

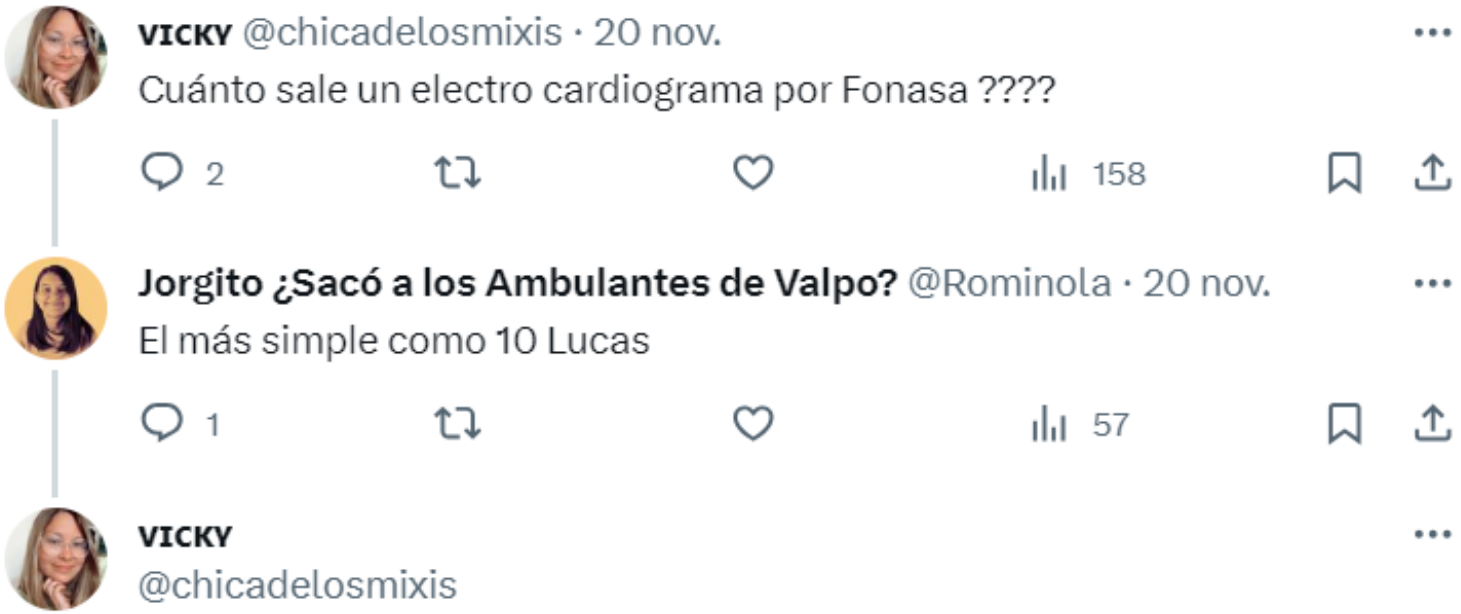
Motivation

FIGURE: *chuta* in Dictionary of Americanisms

Motivation


VICKY @chicadelosmixis · 20 nov.
Cuánto sale un electro cardiograma por Fonasa ????
2
158
Jorgito ¿Sacó a los Ambulantes de Valpo? @Rominola · 20 nov.
El más simple como 10 Lucas
1
57
VICKY
@chicadelosmixis
Chuta y sabes si una orden de hospital la puedo usar en privados???

Motivation



I: How much does an electrocardiogram cost through Fonasa?
R: The simplest one like 10 bucks.
I: **Chuta**, and do you know if a hospital order can be used in private clinics?

Aim of this study

Our aim is to understand what is the meaning of these words, to define them in a way that acknowledges their differences and similarities and predicts their uses.

Research Question

Research Question 1

What is the content expressed by pucha and chuta?

Research Question

Research Question 1

What is the content expressed by pucha and chuta?

1. What is the emotive and epistemic content they express?

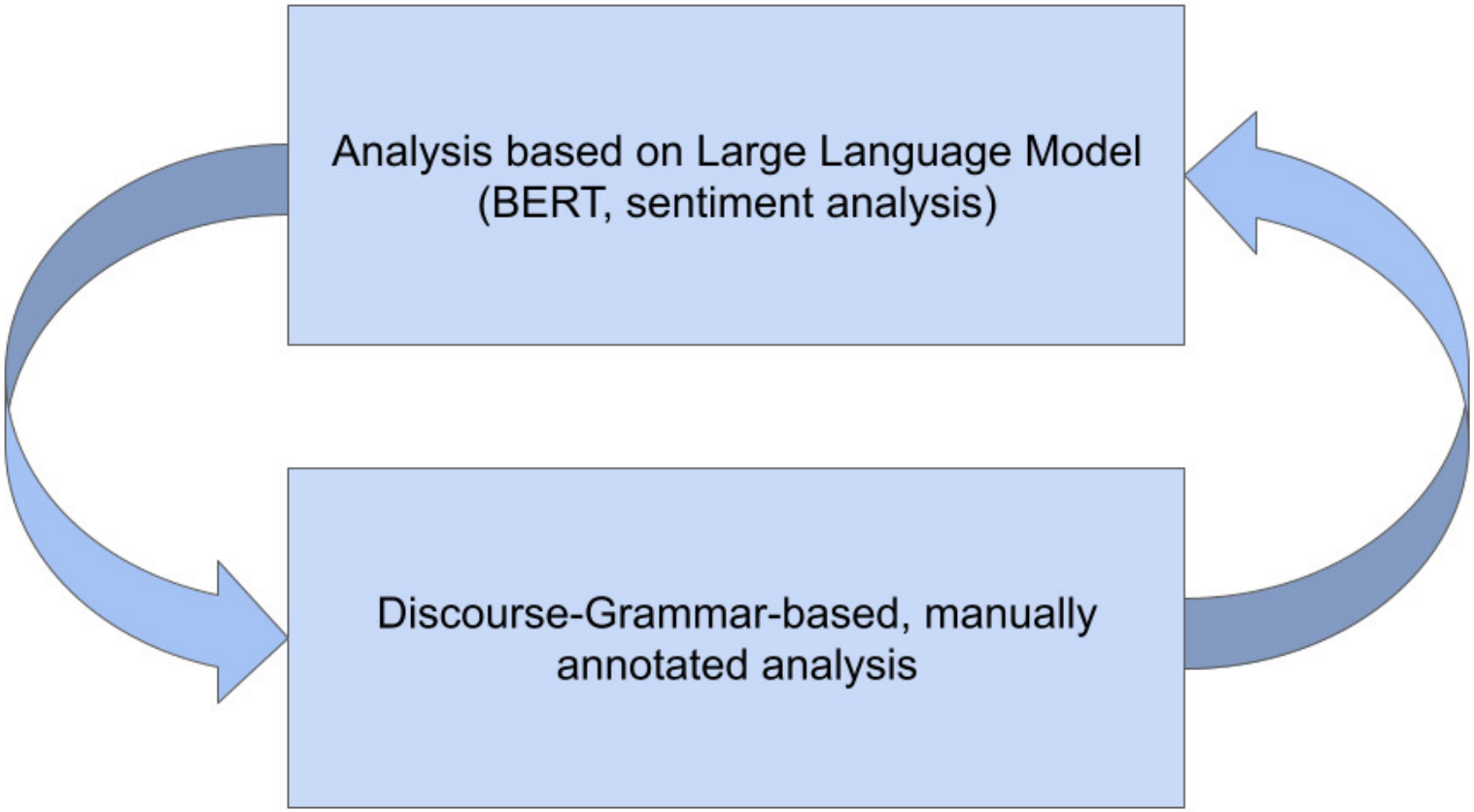
2. How can be these words characterized using representations from LLM?

LLM and contextual-informed annotation

- Interjections are highly contextual. They evaluate information that is part of the common ground. In order to study them, we must collect the interactional context in which they are used.

- We observed the use of these words in interaction creating a corpus from X (aka Twitter). We analyzed the data using manually-annotated labels and computational tools (LLM) that were mutually informed through the process.

Flow Diagram of Methodology



Distribution of interjections

- Both interjections are used primarily as response to an initiation tweet. We observe that even stand-alone utterance either react to some contextual cue or perform a reaction.

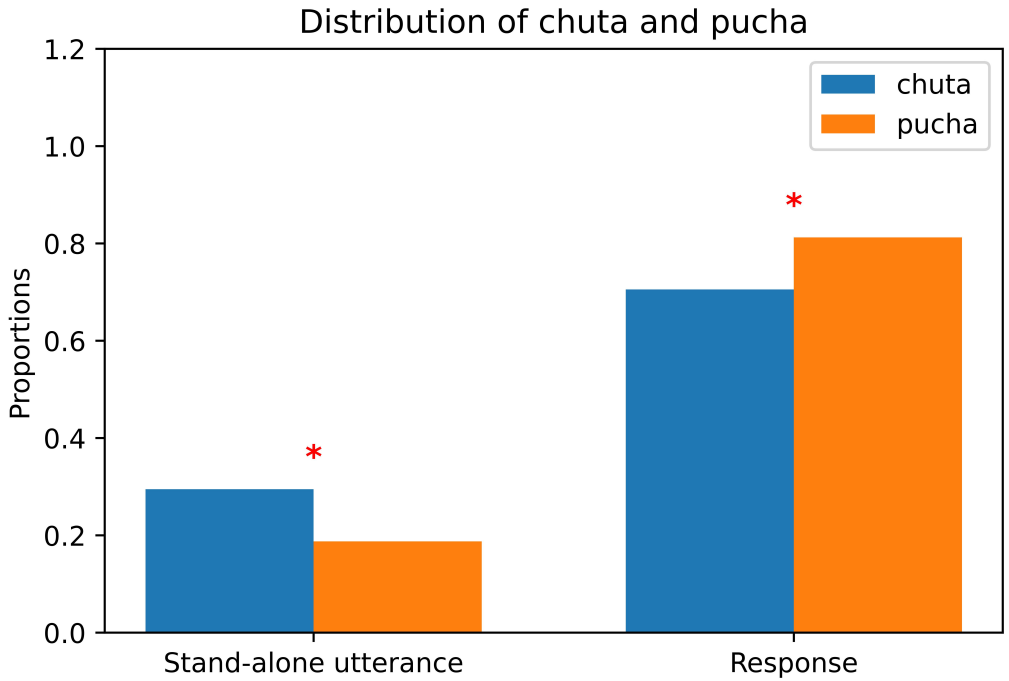
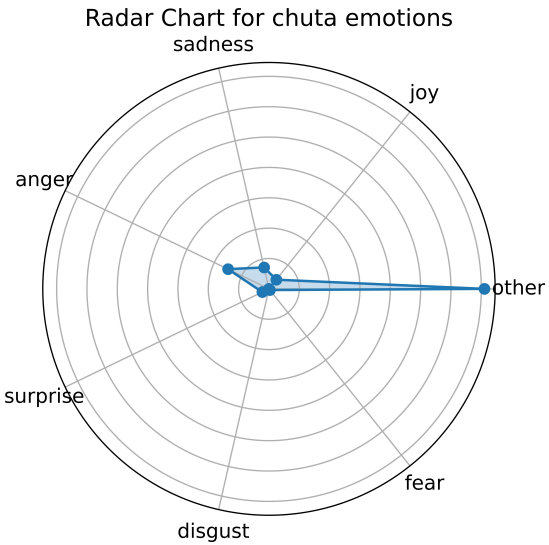
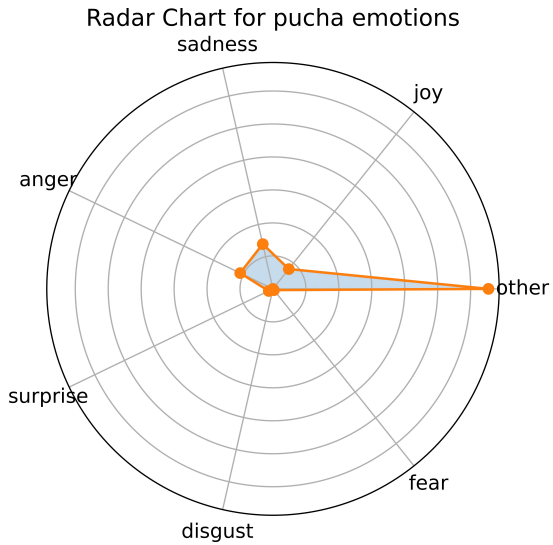


Figure: Chilean Corpus Twitter (2023)

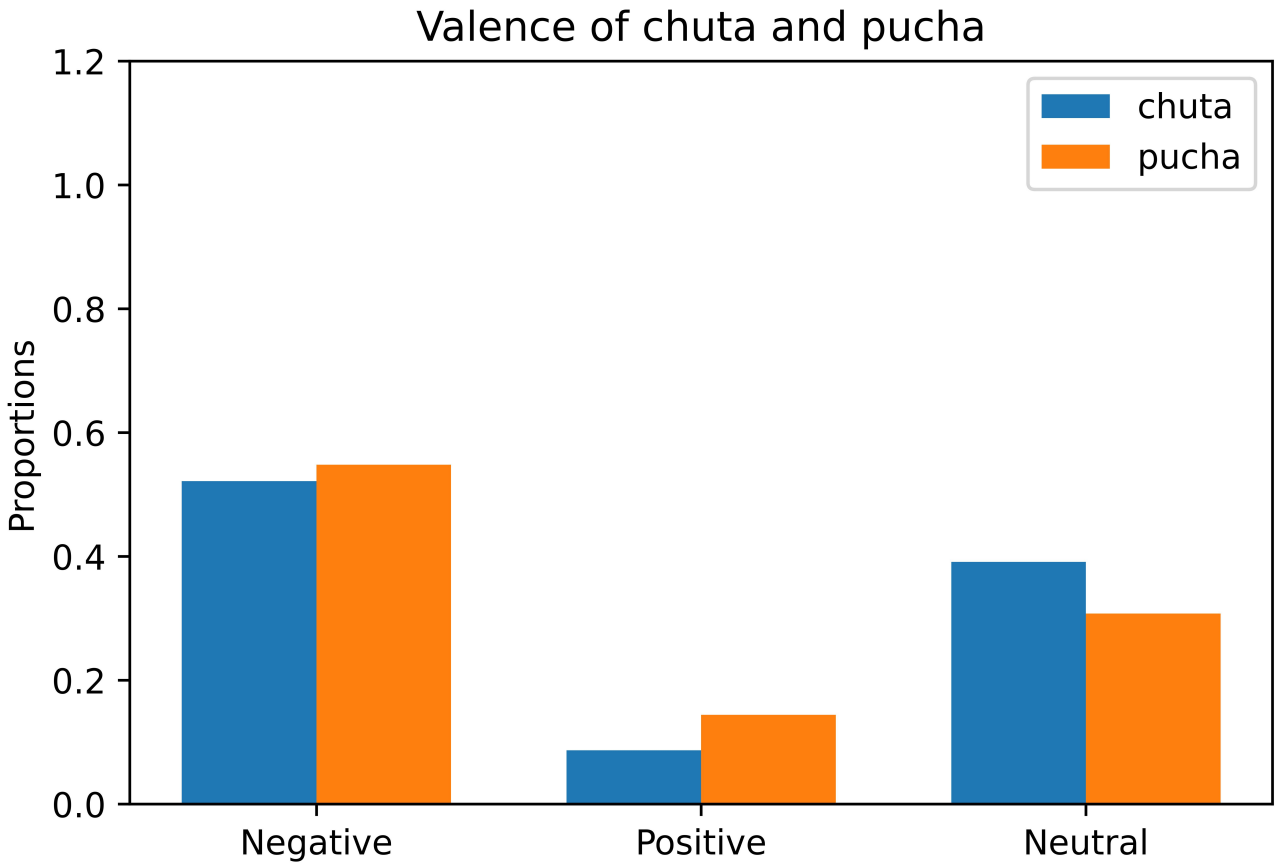
Emotions from Sentiment Analysis



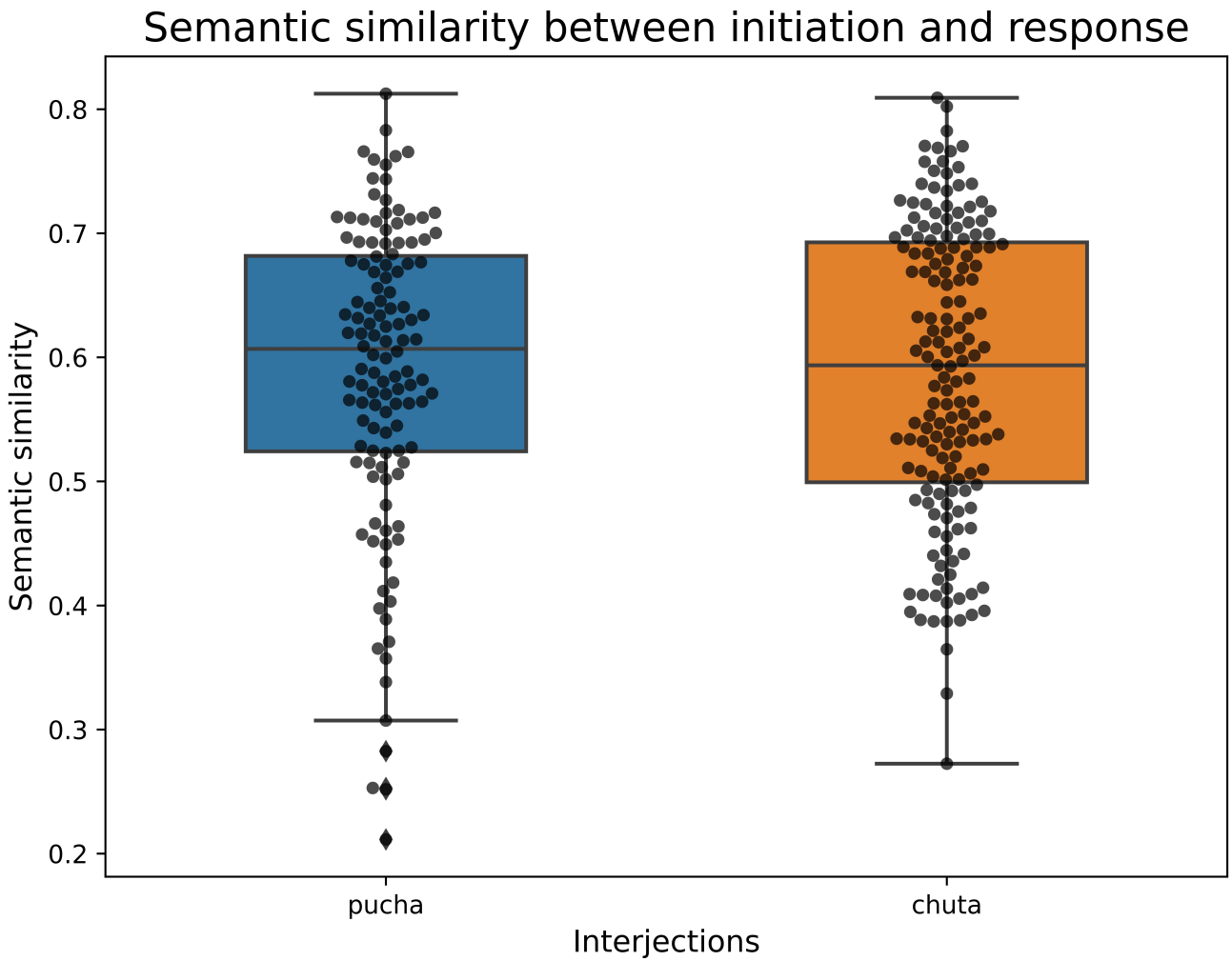


In *chuta* **76.9%** of the tweets are classified as other, followed by anger with **12.9%**. In *pucha* **59.8%** of the tweets are classified as other, followed by sadness with **22.7%** and anger with **9.1%**.

Valence Classification



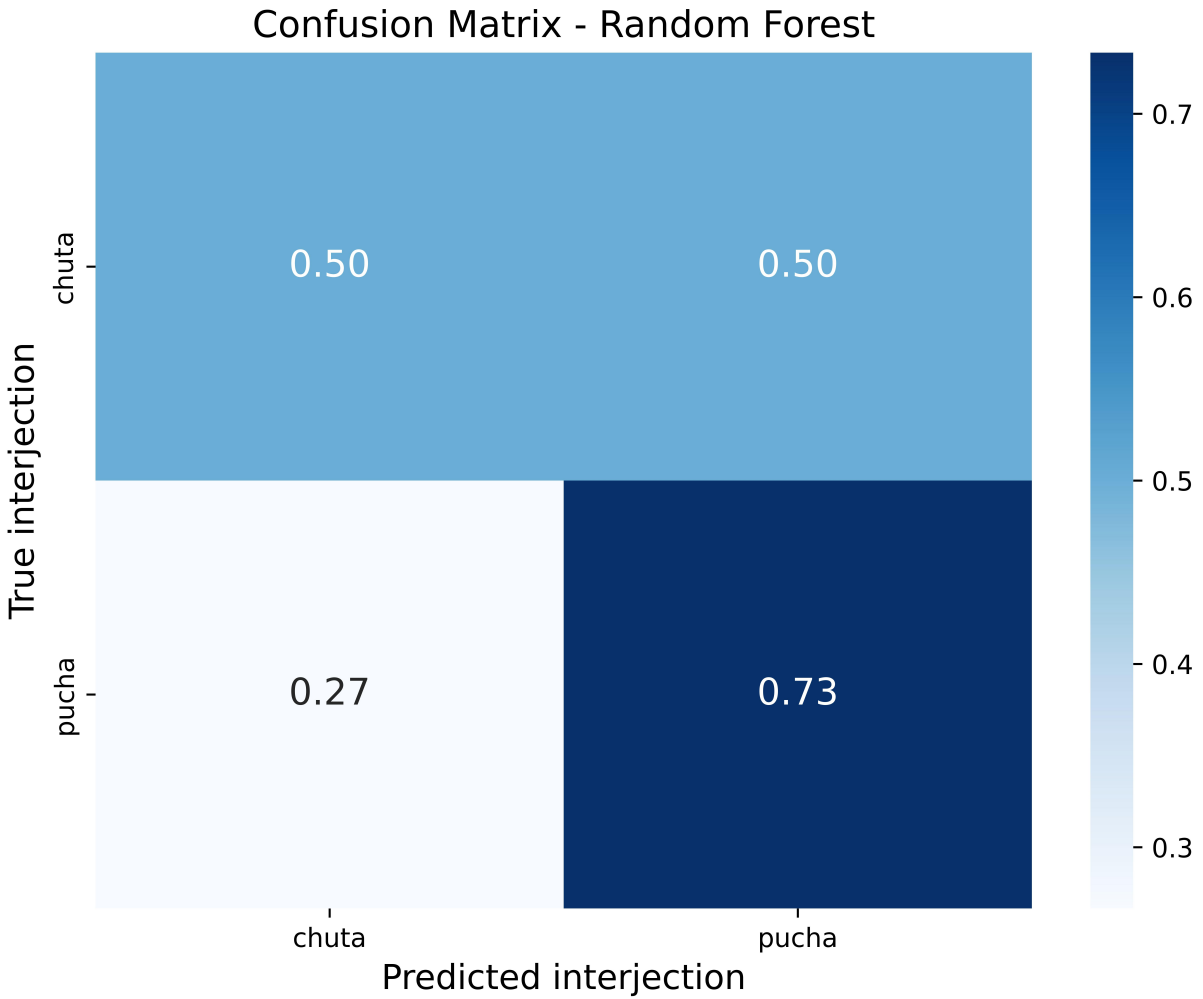
Semantic Similarity Analysis



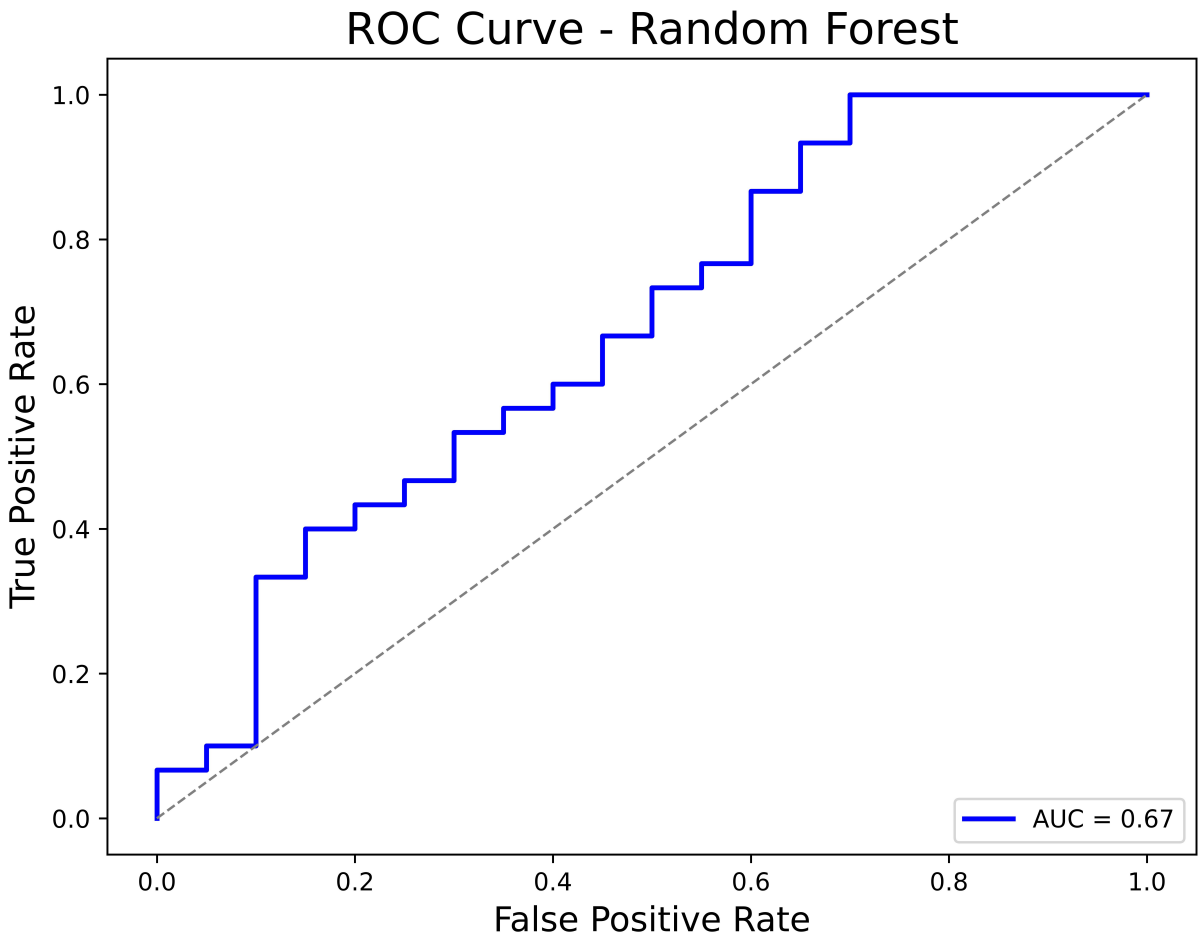
Interjection classification using BERT

We used a BERT model for Spanish to classify collected tweets containing one of both interjections. We used embeddings from BERT and masked both interjections. Then we train a Random Forest model (also Multi-Layer Perceptron (MLP) model) to classify the tweets containing one or the other interjection.

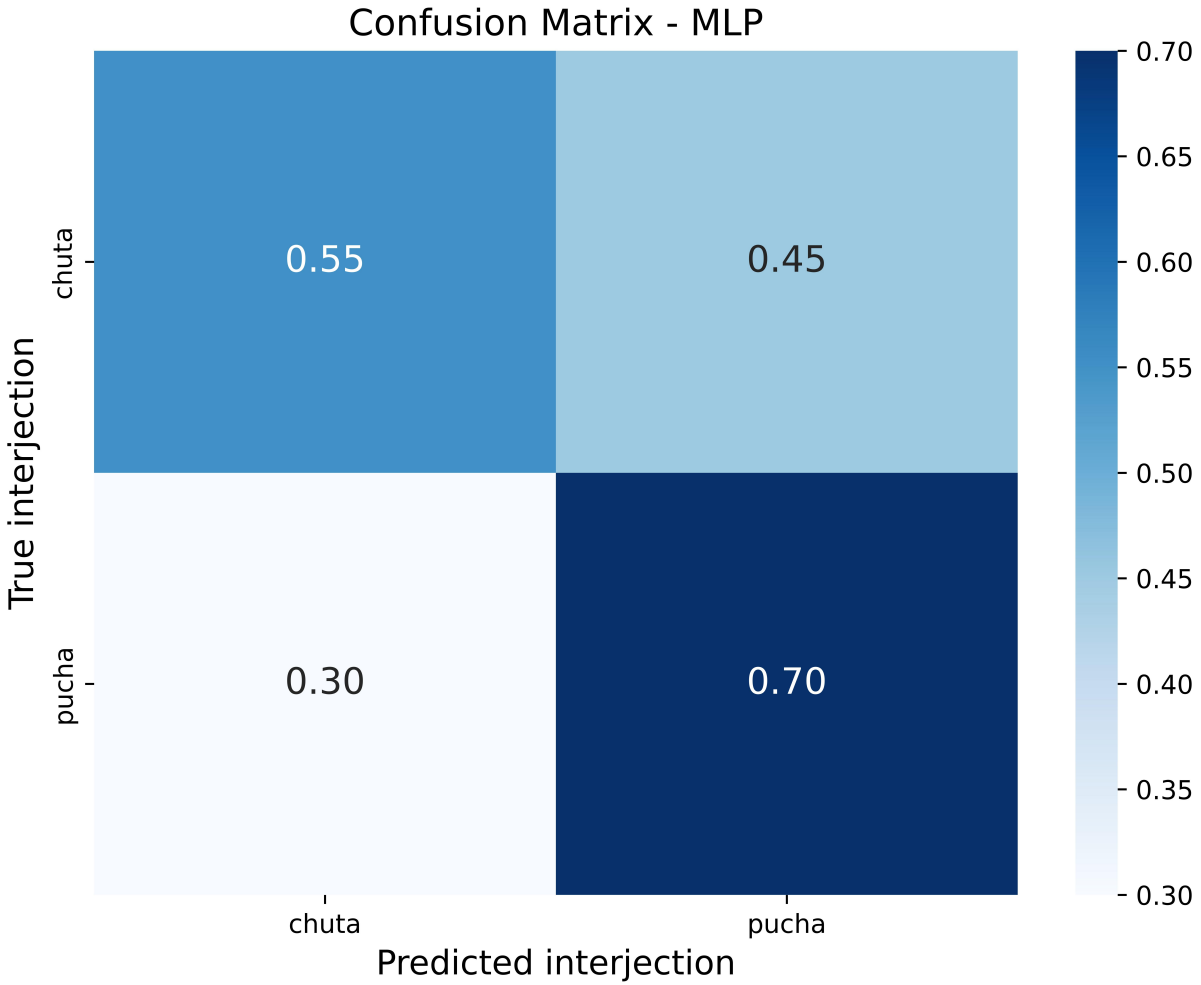
Interjection classification using BERT



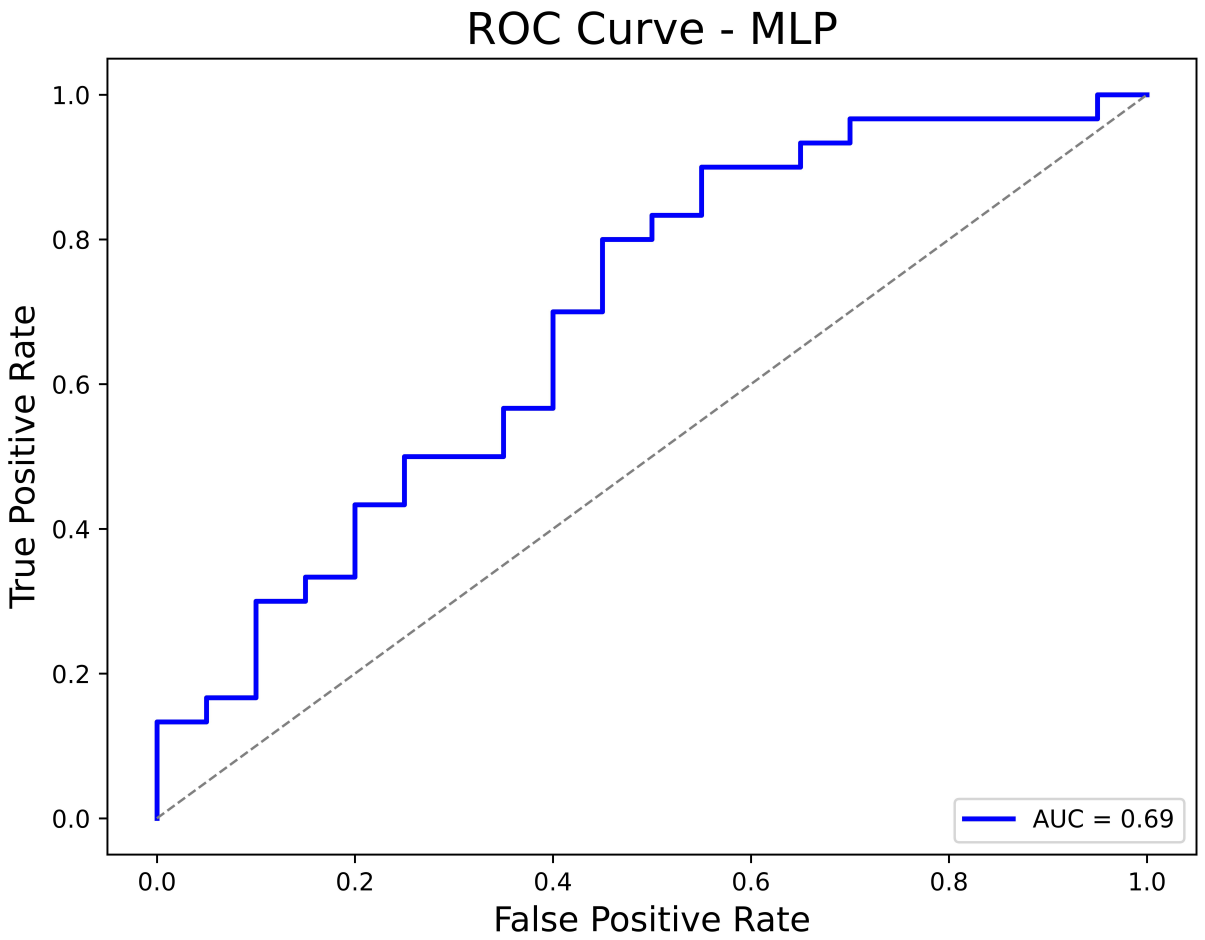
Interjection classification using BERT



Interjection classification using BERT



Interjection classification using BERT



Descriptives of manual annotation

- In both cases, at least 93% of the evaluated propositions are new information to the speaker.

Descriptives of manual annotation

- In both cases, at least 93% of the evaluated propositions are new information to the speaker.

- In 94% of the cases, the truthness of the evaluated proposition is not challenged nor denied.

Descriptives of manual annotation

- In both cases, at least 93% of the evaluated propositions are new information to the speaker.
- In 94% of the cases, the truthness of the evaluated proposition is not challenged nor denied.
- 79% occurrences of chuta and 98% of pucha have negative reading.

Descriptives of manual annotation

- In both cases, at least 93% of the evaluated propositions are new information to the speaker.
- In 94% of the cases, the truthness of the evaluated proposition is not challenged nor denied.
- 79% occurrences of chuta and 98% of pucha have negative reading.
- 98% occurrences of chuta evaluate the proposition as unexpected. Only 21% occurrences of pucha have this reading.

Categorization (from speaker's perspective)

- They have similar distribution.
- Similar emotive value.
- Similar semantic distribution.
- How are they different?

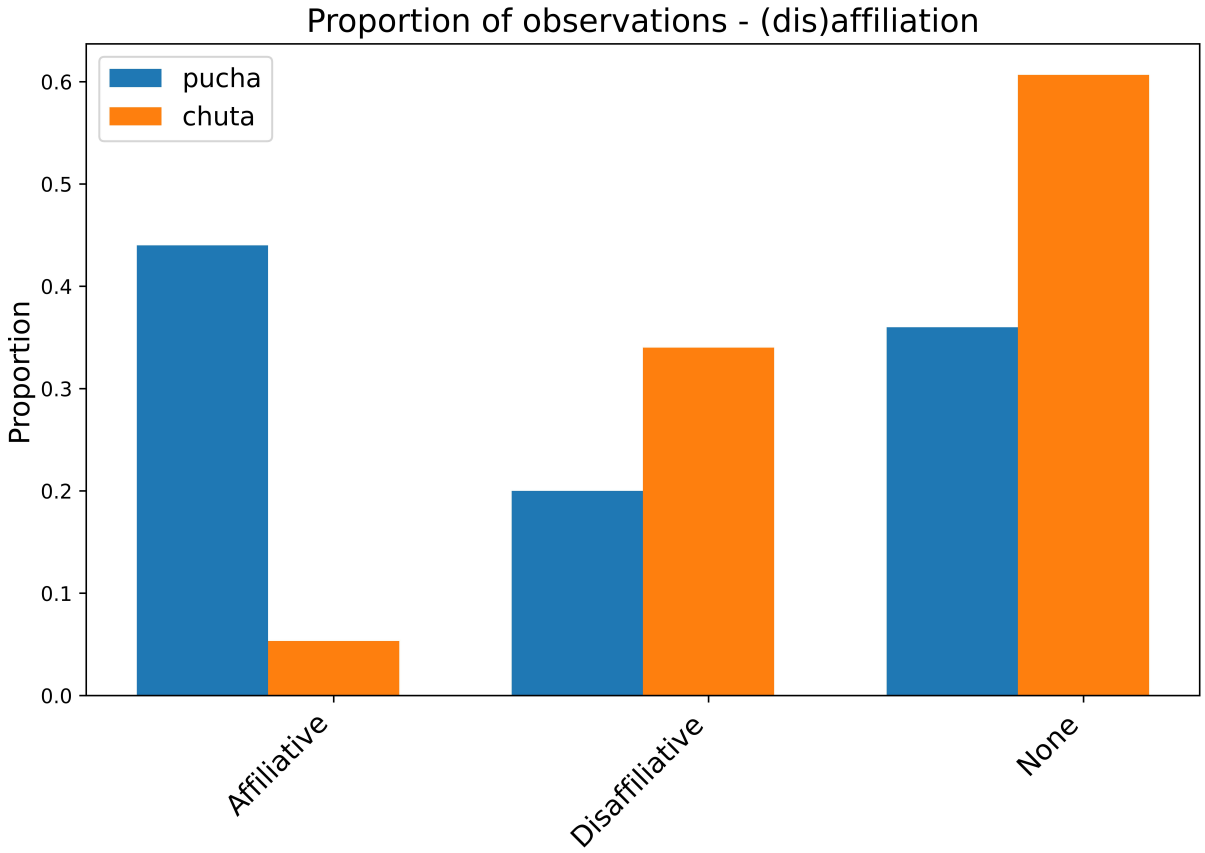
Categorization (from speaker's perspective)

CATEGORY	LABEL		
Action towards addressee's attitude	Affiliative	Disaffiliative	None
Agreement with addressee	Agrees	Disagrees	None
Hedonic valence	Positive	Negative	Neutral
Knowledge	Knows p beforehand	Didn't know p	Unavailable interpretation
Expectedness of p	Expected	Unexpected	Unavailable interpretation
Speaker's beliefs	Believes	Doesn't believe	Uncertain

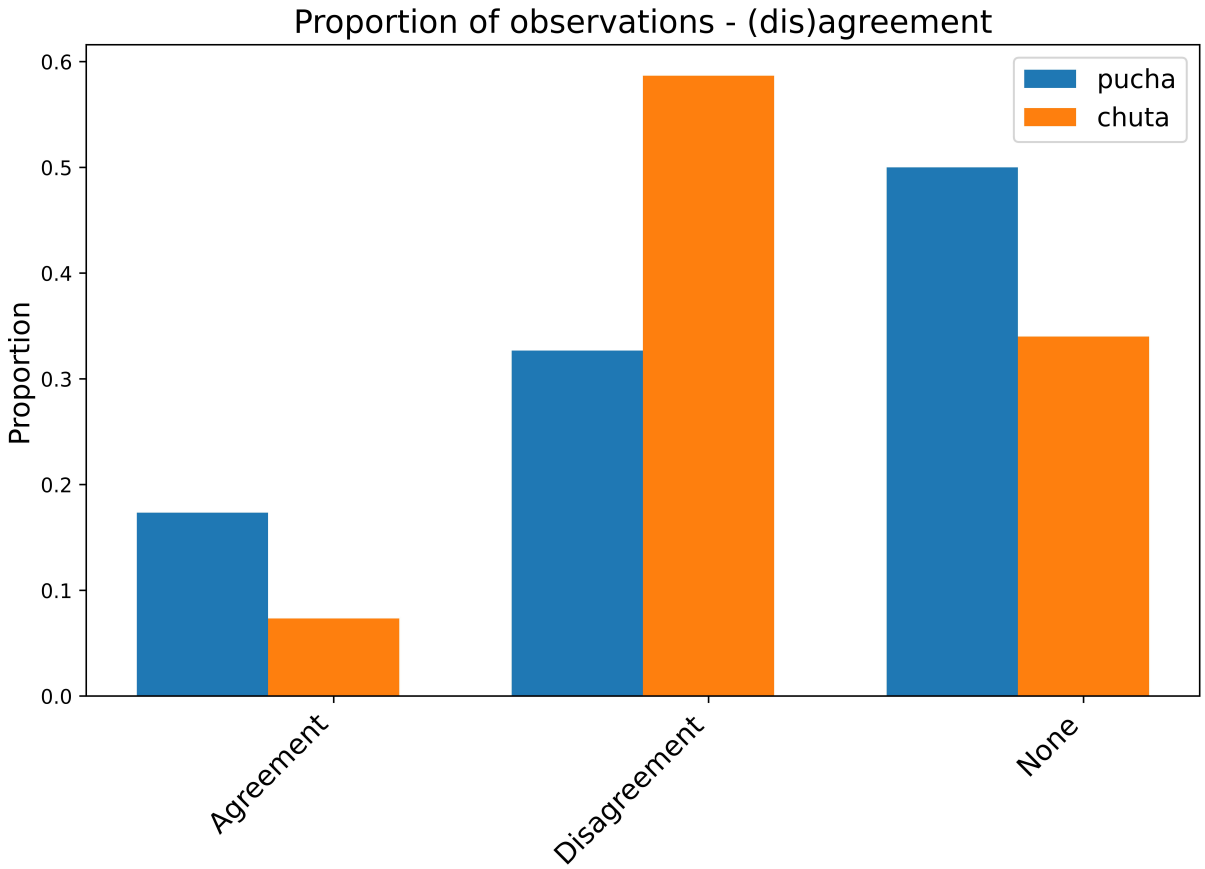
Reference

(Lindström & Sorjonen, 2013)
(Sacks, 1987)
(Feldman-Barret, 2011; Gasper et al., 2019)
(Witschko, 2021)
(Aikhenvaldt, 2012; Mellers et al., 2013)
(Holmberg, 2011/2016)

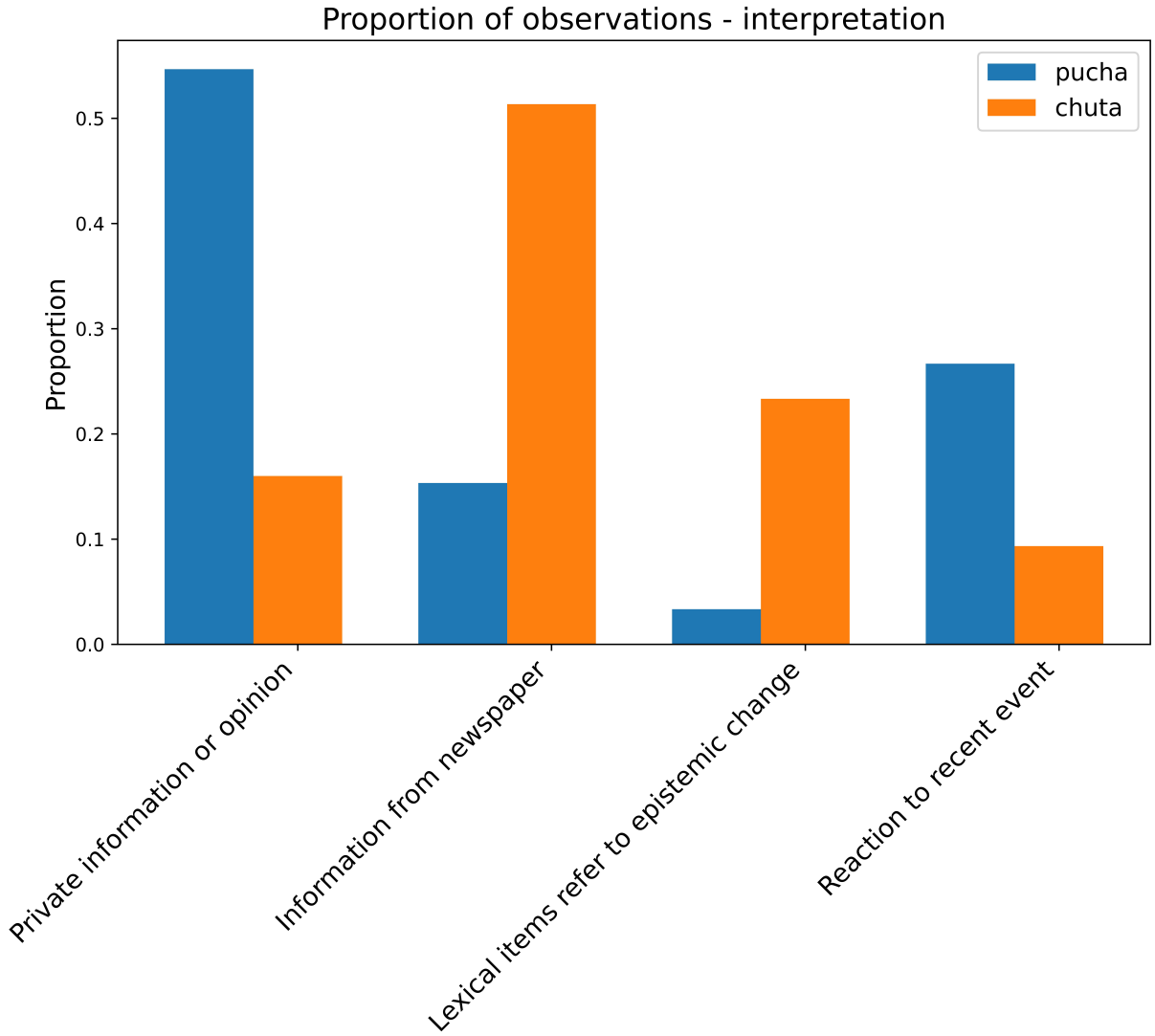
Annotation categories



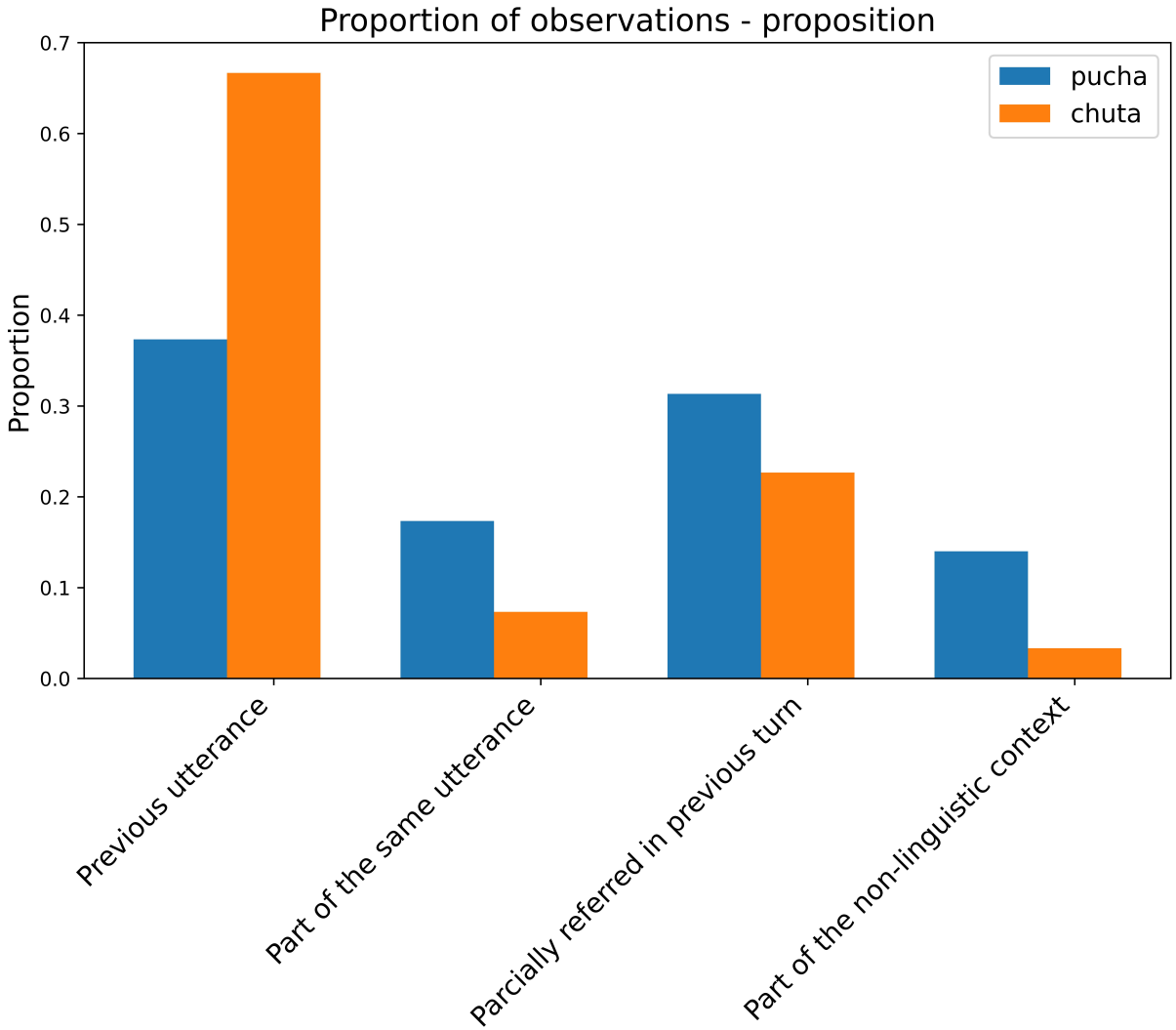
Annotation categories



Annotation categories



Annotation categories



Interjection Dimension Labeling

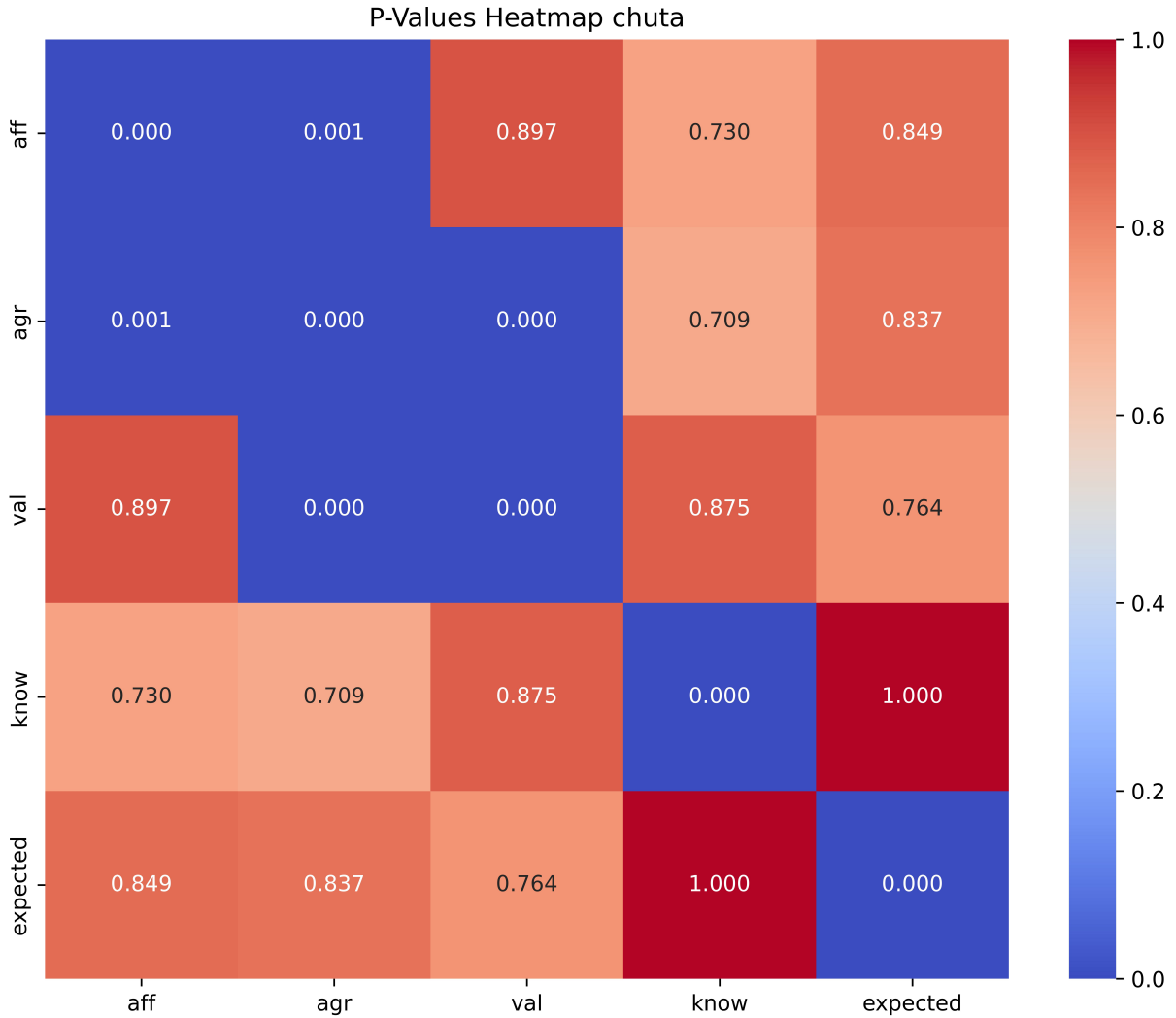


Figure: Heatmap of p-values chi-square test for *chuta*

Interjection Dimension Labeling

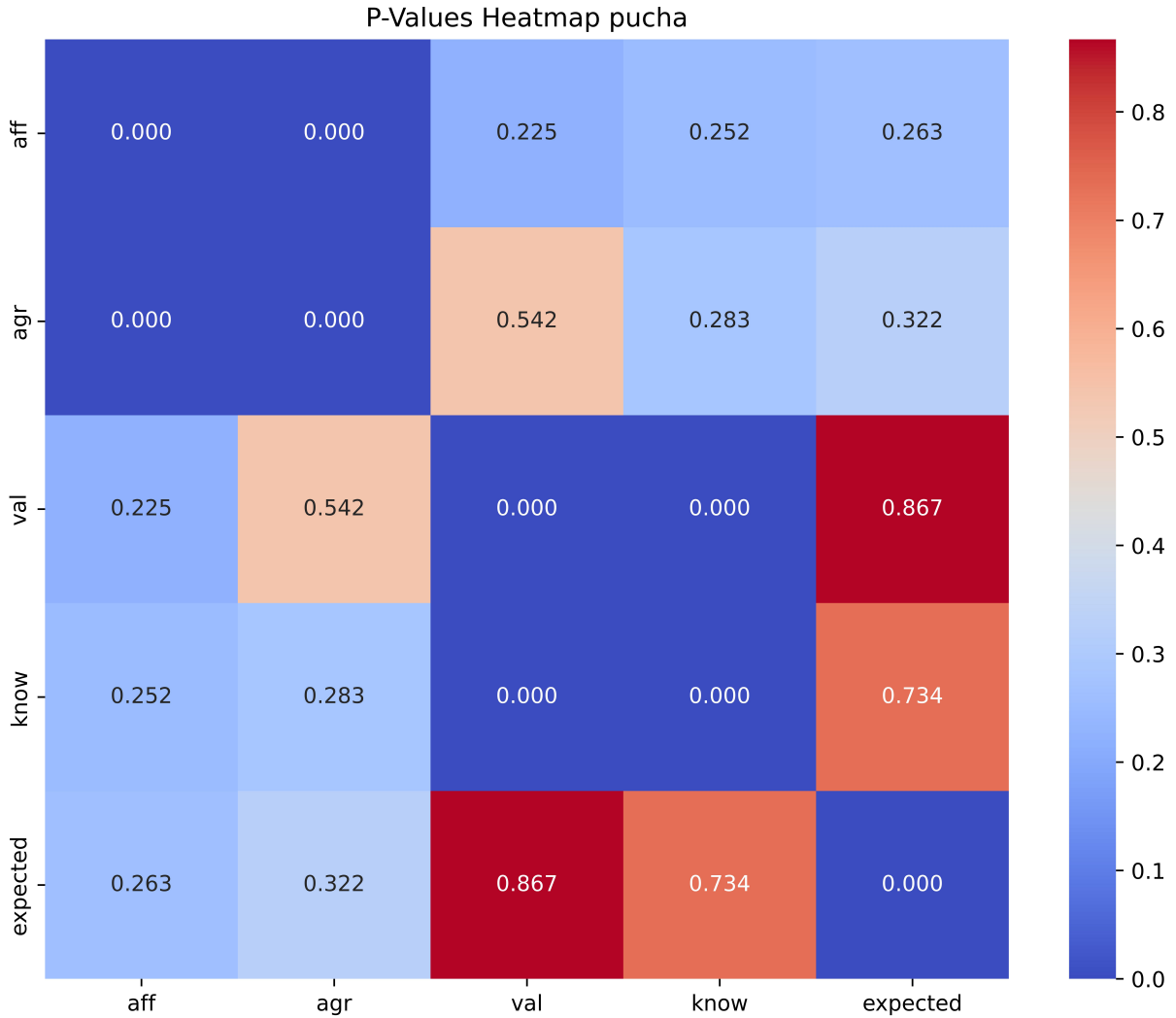


Figure: Heatmap of p-values chi-square test for *pucha*

Insights from the Analysis

- The (automatic and manual) analysis showed that *pucha* tends to be more emotive, while *chuta* tends to be more epistemic. Only the later express an epistemic attitude of surprise.

Insights from the Analysis

- The (automatic and manual) analysis showed that *pucha* tends to be more emotive, while *chuta* tends to be more epistemic. Only the later express an epistemic attitude of surprise.

- This finding correlates with the idea that epistemic states are a necessary condition for the existence of emotive states (Mellers et al., 2013; Rett, 2021)

Discussion

- These interactional particles under study work in the management of the common ground. They serve to acknowledge that the proposition that is shared in interaction was not part of the speaker's ground (beliefs, desires, knowledge) (Wiltschko, 2021).

Discussion

- These interactional particles under study work in the management of the common ground. They serve to acknowledge that the proposition that is shared in interaction was not part of the speaker's ground (beliefs, desires, knowledge) (Wiltschko, 2021).

- Lexical items as expressive interjections contain information regarding affect (a subsystem of emotions) (Feldman-Barrett, 2011/2018) and epistemic states of surprise (Mellers et al., 2013).

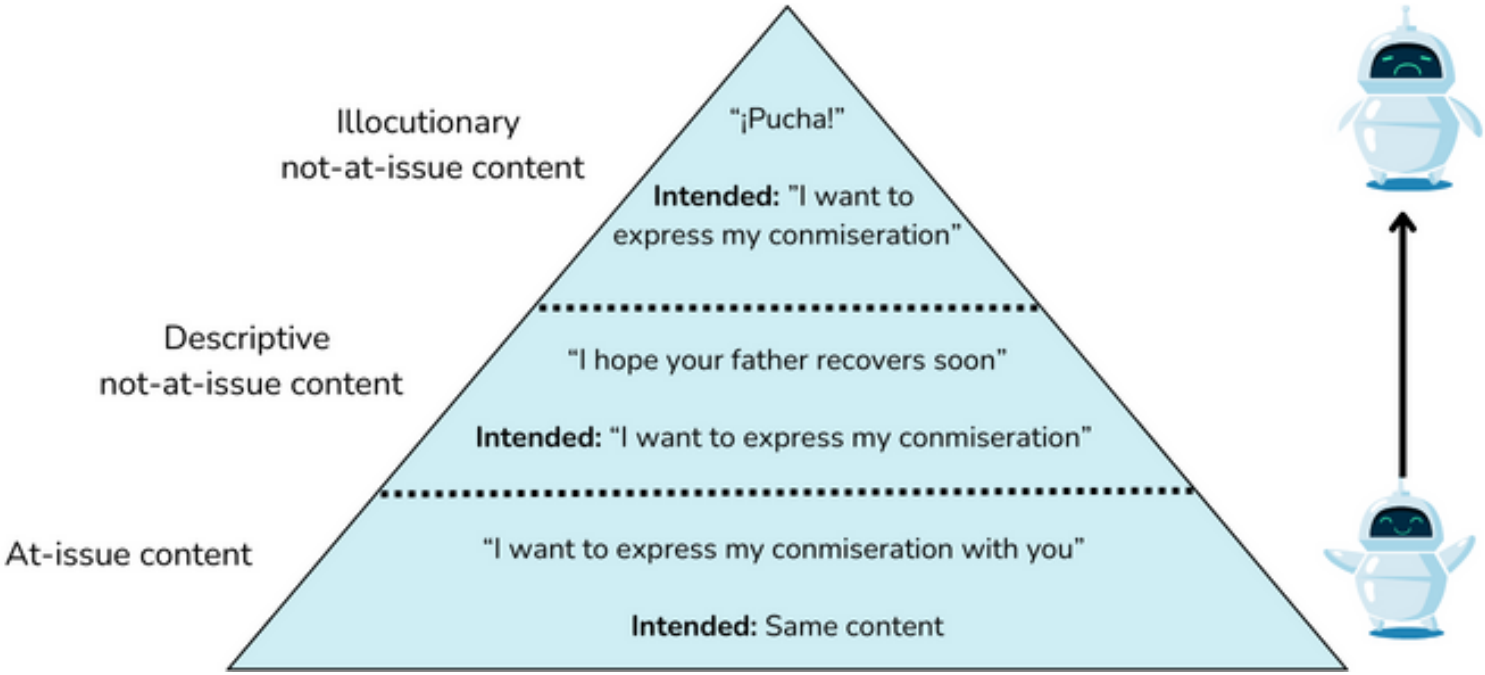
Discussion

- These interactional particles under study work in the management of the common ground. They serve to acknowledge that the proposition that is shared in interaction was not part of the speaker's ground (beliefs, desires, knowledge) (Wiltschko, 2021).

- Lexical items as expressive interjections contain information regarding affect (a subsystem of emotions) (Feldman-Barrett, 2011/2018) and epistemic states of surprise (Mellers et al., 2013).

- Expressive Intejections express culturally specific content that is not textual. This may explain why the manual annotation found differences that the model failed to capture.

Discusion: Levels of semantic meaning



Conclusions

1. Interjections evaluate information that is most of the time elided and it's cultural-specific as part of the Common Ground. For models this evaluation is hard to interpret.

Conclusions

1. Interjections evaluate information that is most of the time elided and it's cultural-specific as part of the Common Ground. For models this evaluation is hard to interpret.

2. Interjections are fundamental to interpret interactants' set of beliefs, and their attitude towards that information.

Conclusions

1. Interjections evaluate information that is most of the time elided and it's cultural-specific as part of the Common Ground. For models this evaluation is hard to interpret.
2. Interjections are fundamental to interpret interactants' set of beliefs, and their attitude towards that information.
3. Computational tools like LLM only can capture part of the (epistemic, emotive) meaning, specially when high contextual cues are involved.

Conclusions

1. Interjections evaluate information that is most of the time elided and it's cultural-specific as part of the Common Ground. For models this evaluation is hard to interpret.
2. Interjections are fundamental to interpret interactants' set of beliefs, and their attitude towards that information.
3. Computational tools like LLM only can capture part of the (epistemic, emotive) meaning, specially when high contextual cues are involved.
4. The methodology is applicable to any other language, as the labels are motivated by cross-linguistic evidence. Nonetheless, the analysis of such labels must be cultural-specific.

Limitations

Recent restrictions to web scraping in X (a.k.a. Twitter).

Limitations

Recent restrictions to web scraping in X (a.k.a. Twitter).

Future directions

1. How do other interjections work?
   - Chilean-Spanish interjections.
   - Interjections in other languages.
   - Cross-linguistic analysis.

Future directions

1. How do other interjections work?
   - Chilean-Spanish interjections.
   - Interjections in other languages.
   - Cross-linguistic analysis.

2. Performance of prosody through vowel lengthening (*puuuucha*) and intonation contour.

Future directions

1. How do other interjections work?
   - Chilean-Spanish interjections.
   - Interjections in other languages.
   - Cross-linguistic analysis.

2. Performance of prosody through vowel lengthening (*puuuucha*) and intonation contour.

3. Use of generative models (like GPT) to observe how they would use different interjections.

Future directions

1. How do other interjections work?
   - Chilean-Spanish interjections.
   - Interjections in other languages.
   - Cross-linguistic analysis.

2. Performance of prosody through vowel lengthening (*puuuucha*) and intonation contour.

3. Use of generative models (like GPT) to observe how they would use different interjections.

4. Use linguistic data to model possible hierarchies between epistemic states and emotive states.

Thank You

Thank you for your attention.

Acknowledgements

This work was funded by the National Agency of Research and Development of Chile (ANID) / DOCTORADO BECAS CHILE/2022 - 72220332 - 72220098, Pompeu Fabra University