

# The Discriminant Power of Interjections in Forensic Speaker Identification

Sinjini Chakraborty<sup>1</sup>, Grace Suneetha Didla<sup>2</sup>

The English and Foreign Languages University, Department of Phonetics, Hyderabad, India

The English and Foreign Languages University, Department of Phonetics, Hyderabad, India

**Abstract:** *While interjections have been looked at by Traditional Grammarians as an emotional language, other researchers have argued that interjections are deictic items (Poggi, 2008) which are essentially verbal tokens of the speaker's emotional state "in praesentia" (Hockett, 1960). Interjections have been categorised in several ways but all interjections can be said to be a codified signal that is stored in our long-term memory as emotion-signal tokens. These tokens are "neither universal nor meaningless. On the contrary, they are language-specific, and they are meaningful" (Wierzbicka, 2003). Interjections are spontaneous utterances that are deeply rooted in the verbal repertoire of the speakers, and can thereby, be identifying markers of cultures and individuals. In that light, interjections hold relevance to Forensic Speaker Identification. The primary objective of this study was to identify whether interjection usage is influenced more by (a) linguistic background, (b) geographical region, or (c) idiosyncratic factors. To this effect, interjections for 9 different emotions were looked at (a) in Arabic, as spoken across 3 different countries - Lebanon, Iraq and Palestine and (b) in India, across 3 different languages - Telugu, Bengali and Malayalam. The outcome of the study suggests that interjection usage is influenced more by linguistic commonality than geographic belonging. However, the contribution of geographic factors cannot be ruled out.*

**Keywords:** Interjections, emotions, speaker profiling, Forensic Speaker Identification, LADO

## 1. Introduction

Interjections have been looked at by Traditional Grammarians as an emotional language [1]. Sapir claimed that "interjections are among the least important of speech elements", and that "they are never more, at best, than a decorative edging to the ample complex fabric" [2]. However, other researchers have argued that interjections are "neither universal nor meaningless. On the contrary, they are language-specific, and they are meaningful" [3]. Poggi states that interjections are deictic items which are essentially "holophrastic" verbal tokens of the speaker's emotional state "in praesentia" [1], [4]. In keeping with this, Ameka & Wilkins have defined interjections as "context-bound linguistic signs" that "index elements in the extra-linguistic context", and thus, are a sub-class of "situation-bound utterances (SBUs)" [5].

In addition to its definition, there has been much confusion around establishing a category in language that represents the unique features of interjections. But all interjections have been agreed upon to be a codified signal that is stored in our long term memory as emotion-signal tokens. Several researchers have worked on interjections in multiple languages in an attempt to categorise them (e.g.: [1], [6]-[9]). On the basis of their lexical structure, interjections have been categorised as (a) primary interjections, which do not belong to any other word-class in the lexicon, and (b) secondary interjections, which "have an independent semantic value" and can be "used conventionally as utterances by themselves to express a mental attitude or state" [10].

Poggi introduces another classification of interjections on the basis of their lexical structure: (a) univocal interjections are those that have a single meaning irrespective of the context it is used in, while (b) plurivocal interjections could

have several meanings depending on the context used in, or intonation and other such lexical markers [1]. There have been discussions around whether interjections could simply be classified as an exclamation or an onomatopoeic item.

But Wierzbicka maintains that interjections are laden with meaning, and offers a semantic classification of interjections into three categories: (a) emotive, (b) volitive and (c) cognitive interjections [11]. While emotive interjections express a feeling or an emotion ('Ouch!' to express 'pain'), volitive interjections express the speaker's call for action ('Shh!' to ask for silence), and cognitive interjections indicate confirmation responses from the speaker ('Aha!' to indicate understanding something). Ameka & Wilkins provide a pragmatic classification of interjections in terms of communicative functions into three categories: (a) expressive, which includes both emotive and cognitive interjections, (b) conative, which overlaps with the semantic category of volitive interjections, and (c) phatic interjections, which are essentially conversation regulators [5].

This paper specifically looks at a sub-classification of interjections that performs the communicative functions of 'expressive' interjections, but can be better described as "specialised emotive interjections". Mao defines this category of interjections as primary interjections that typically express certain specific emotions and "are language-specific, and hence need acquiring in the second language learning, especially in that of spoken language" [12]. Additionally, it has been documented that some interjections are learnt very early by native speakers, however, they can be quite difficult for non-native speakers to acquire [8]. Drawing from these, one could assume that interjections are essentially spontaneous utterances that are deeply rooted in a native speaker's verbal repertoire, and are laden with both 'social intent' and 'communicative intent' [13]. Interjections, thereby, can function as identifying

markers of one's culture and thus, hold the possibility of assisting in forensic speaker identification, specifically in the domain of Language Analysis for Determination of Origin (LADO).

## 2. Objectives of the Study

This study attempts to:

- Explore whether the usage of interjections is influenced more by linguistic, geographical, or idiosyncratic factors,
- Identify the patterns in interjection usage for 9 different emotions across 3 countries (Lebanon, Iraq, Palestine) that speak the same language (Arabic), and across 3 different languages (Telugu, Bengali, Malayalam) that are spoken in the same country (India),
- Offer an account of the general terminology of interjections used across languages and countries.

## 3. Scope of the Study

The scope of this study is limited to observing interjection usage:

- Across three languages from India, a country that speaks more than 120 odd languages [14], and
- Arabic, as spoken in three countries, out of the 26 countries where Arabic is officially recognized.

This study looks at the following 9 different emotions, all of which belong to a specific category of interjections, namely, 'specialised emotive interjections':

Appreciation	Happiness	Surprise
Disgust	Pain	Sympathy
Fear	Regret	Relaxation

## 4. Methodology

### 4.1 Subjects

This study included a total of 18 participants; 10 female and 8 male subjects. The age of the participants varied from 20-60 years. The data was collected from university students and teaching faculty at The English and Foreign Languages University, Hyderabad. The minimum qualification of the participants was graduation. All the participants were native speakers of the languages that they have provided data for, and were at least basic users of English. Only 3 of the 18 participants were trained in Phonetics. None of the participants exhibited any language disability.

For the same-language (Arabic) data, 3 subjects each from the countries Lebanon, Iraq and Palestine were included in the study. All the Arabic-speaking participants were, at the time, enrolled in a Basic English course at the university. The data from Arabic-speaking participants was collected in batches (3 batches in total, of 3 participants each) of individuals belonging to the same country, and who spoke the same language - Arabic.

For the same-country (India) data, 3 native speakers each of the languages Telugu, Bengali and Malayalam, were included in the study. All the participants in the Indian data group were proficient users of English. The data from the

Indian participants were collected in batches of individuals who spoke the same native tongue (a total of 3 batches with 3 participants each).

The participants were told that the study was on interjections, but they were not told the specific intent of it to control for any data bias that could be introduced. All the participants signed a consent form to provide data for this study prior to data collection.

### 4.2 Method

The participants were first briefed on the area of the study, and then told how they were to perform the tasks. A questionnaire in English, developed for this study, was introduced to the participants. All doubts regarding either the tasks or the questionnaire were satisfactorily clarified before the task was attempted by the participants. The questionnaire required the participants to fill in their personal details, and then attempt two 'fill in the blank with interjections' exercises. Both the exercises attempted at eliciting interjections for the same set of 9 different emotions.

As part of the first exercise, the participants were required to read a set of 9 contextualised situations targeting each of the 9 different emotions. A few examples from the questionnaire have been given in Table 1.

**Table 1:** Question samples from situational category

1	An artist painted five beautiful pictures in less than 20 minutes. How would you show your appreciation?
2	You step on your neighbour's dog poop. Show your disgust.
3	You see a tiny puppy with an injured leg. How do you show sympathy?
4	You just took a cold shower on a hot afternoon. How would you express relaxation?

This exercise was aimed at eliciting spontaneous responses. The participants were asked to first, verbally respond to, and then write the same response down on the questionnaire in English. Help was provided to basic users of English to write down their interjections in English letters. This category will henceforth, be referred to as the situational category.

The second exercise aimed at eliciting interjections in the absence of any contextual clue. The participants were given the same set of 9 emotions in English and asked to respond to the same in their native language (see Table 2).

**Table 2:** Question samples from generic category

When...	In English, I would say...	In <u>my mother tongue</u> , I would say...
1. I appreciate something	Wow!	
2. I'm disgusted	Yuck!	
3. I feel sympathy	Tsk Tsk!	
4. I feel relaxed	Ah!	

Care was taken to ensure that all the participants were familiar with the English interjections. This exercise attempts at extracting normative interjections as used by speakers of the languages included in this study. This will be referred to as the generic category, as these will indicate

whether the target language has any formal interjection in its repertoire for these specific emotions. Again, the participants were asked to first verbally respond, and then transliterate their responses in English and note those down in the questionnaire. Data for both the categories were collected in a single session, in the order of the exercises. Each session lasted for 15-20 minutes.

### 4.3 Analysis

On completion of data collection, all the responses for each of those 9 different emotions were noted down, separately for the Arabic and Indian data, from the questionnaires. The number of matches and mismatches in responses of participants (a) within and across countries, (b) within and across languages, (c) within and across categories and (d) within and across emotions were tabulated. The analysis of the data was then carried out to answer the following questions:

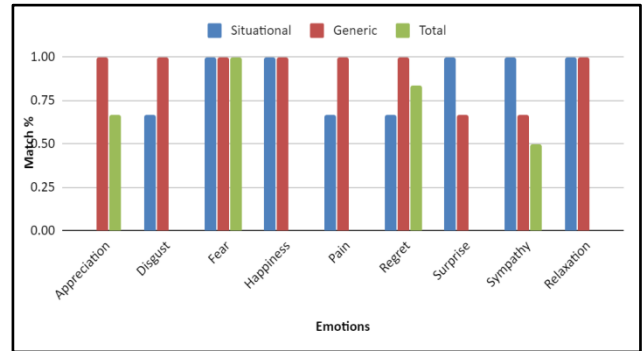
- What are the similarities and dissimilarities in interjection usage both within and across the Arabic-speaking countries (Lebanon, Iraq and Palestine)?
- What are the similarities and dissimilarities in interjection usage both within and across the 3 languages (Telugu, Bengali and Malayalam) from India?
- What are the noticeable/significant patterns in interjection usage across the 9 different emotions?
- Is there any common pattern in the terminology used for interjections across the languages and countries taken into this study?

## 5. Observations

### 5.1 Interjection usage within each Arabic-speaking country

This part of the experiment takes into account only one language, Arabic, as spoken in 3 different countries, namely - Lebanon, Iraq and Palestine. The maximum number of responses collected were: 54 per country (18 per participant); 27 per category (9 per participant); 18 per emotion (2 per participant). A total of 4 responses were not submitted.

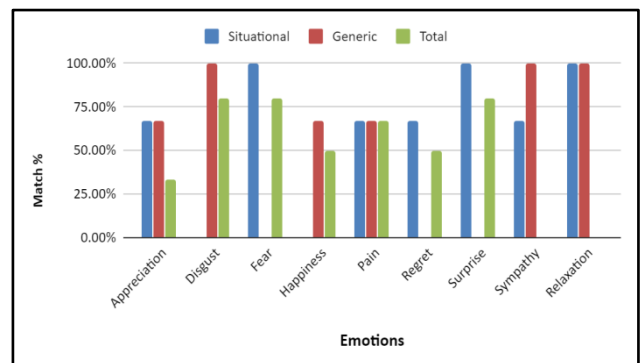
- Lebanon displays a peculiar pattern in interjection usage: wherein, the majority of the emotions display a 100% match in the generic category, even when there aren't considerable matches in the corresponding situational category. It must also be noted that Lebanon displays a maximum percentage of matches within both the situational and the generic category of data (see Fig. 1(a)). Responses for the situational category show a high percentage of matches for the majority of the emotions ('fear', 'happiness', 'surprise', 'sympathy' and 'relaxation'); the exception being for the emotion 'appreciation', wherein it displays no match at all. Similarly, responses for the generic category show a high percentage of matches; in fact, a 100% match for the majority of the emotions ('appreciation', 'disgust', 'fear', 'happiness' and so on).



**Figure 1(a):** Comparison of matches in interjection usage within Lebanon

It may be noted here that the responses from the participants for the emotions 'fear' and 'regret' were essentially the same across the two categories (see Appendix\_1); the difference being in spelling and morphological variations. However, matches across the categories within Lebanon seem quite scanty; the majority of the emotions show a 0% match across categories.

- From an overall view, Iraq seems to have more inter-speaker variations within the country, with the participants differing in their responses and the majority of the emotions having less than 60% matches both within and across the situational and the generic categories (see Fig. 1(b)). The data from Iraq displays a 100% match in interjection usage within a category for only the emotions 'fear', 'surprise' and 'relaxation' (within the situational category), and 'disgust', 'sympathy' and 'relaxation' (within the generic category).



**Figure 1(b):** Comparison of matches in interjection usage within Iraq

However, on the contrary, matches across categories in the data from Iraq seem to be the highest amongst the three Arabic-speaking countries included in this study. It might be interesting to note that in the case of the emotion 'appreciation', the responses from the participants included groups of similar phrases that were used across categories (see Appendix\_1).

- It can also be noted in Appendix\_1 that the Palestine group uses interjections that were remarkably consistent within the two categories of interjections. Generic tokens for the majority of the emotions ('appreciation', 'disgust', 'fear', 'happiness', 'regret' and 'sympathy') show a 100% match (see fig. 1(c)). Similarly, all the participants from Palestine mostly used the exact same words/phrases as

interjections for most of the emotions (like ‘appreciation’, ‘fear’, ‘happiness’, ‘pain’ and so on). This makes Palestine the group that shows the least inter-speaker variability in interjection usage.

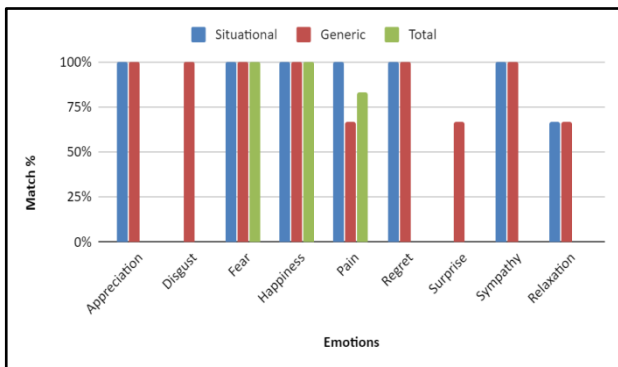


Figure 1(c): Comparison of matches in interjection usage within Palestine

Quite surprisingly though, the responses of the participants from Palestine show the least matches in interjection usage for a specific emotion across the two categories of data. The only exceptions were the responses for ‘fear’ and ‘happiness’, wherein a 100% match occurred. It may be also noted here that these are the only two instances in the Arabic data that have a 100% match across categories within a country.

- Appendix\_1 also shows that participants from both Lebanon and Palestine display a consistent usage of interjections within each category. Owing to this, Lebanon and Palestine showcase distinct intra-cultural majority patterns in interjection usage for both the categories (see figs. 1(a), 1(c)).
- Generic interjections show a greater match percentage consistently within each Arabic-speaking country included in this study. This is probably indicative of the fact that normative interjections exist, and are popular in speech for these emotions in these Arabic-speaking countries. Instances can be found in the data for Palestine and Iraq (figs. 1(b) and 1(c)). Matches in interjection usage across the situational and generic categories are significantly low in the data from all the Arabic-speaking countries, with very few instances of a 100% match in token usage: for the emotion ‘fear’ in Lebanon, ‘fear’ and ‘happiness’ in Palestine and none in Iraq.

5.2 Interjection usage within each language from India

This part of the experiment takes into account 3 Indian languages, namely - Bengali, Telugu and Malayalam. The maximum number of responses collected were: 54 per language (18 per participant); 27 per category (9 per participant); 18 per emotion (2 per participant). A total of 9 responses were not submitted.

- As is represented in Fig. 2(a), data from Telugu speakers show a consistent usage of interjections in the situational tokens: more than half of the emotions show a 100% match and the remaining emotions too, show more than 50% match in interjection usage. On the contrary, matches in the generic category are completely absent in two of the emotions (‘happiness’ and ‘surprise’). But for the rest of the emotions, the generic category shows more than 50%

matches, consistently.

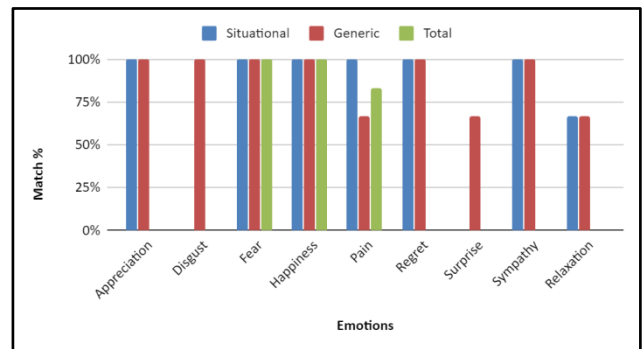


Figure 2(a): Comparison of matches in interjection usage across Telugu speakers

As can be seen in Appendix\_2, Telugu data displays that more than one primary interjection has been used interchangeably for a particular emotion in a lot of instances, to express similar feelings (see ‘disgust’ and ‘regret’).

- As for the Bengali data, there is a peculiar trend in the usage of the interjections wherein most of the interjections across emotions is some allomorph of invocations to God or one’s father/mother (see Appendix\_2: ‘orey baba re’, ‘haayebhogobaan’, ‘o maa go’ and so on).

From a distant view, data representations for each Indian language seem quite consistent with one another. In the Bengali data, match percentages seem to be the most consistent: always 50% or above. This could be a consequence of the interjections being allomorphs of each other, and used interchangeably for various emotions with the differences set in their intonation. The only exception being for the emotion ‘happiness’ for which all the languages seem to have a peculiar pattern in interjection usage.

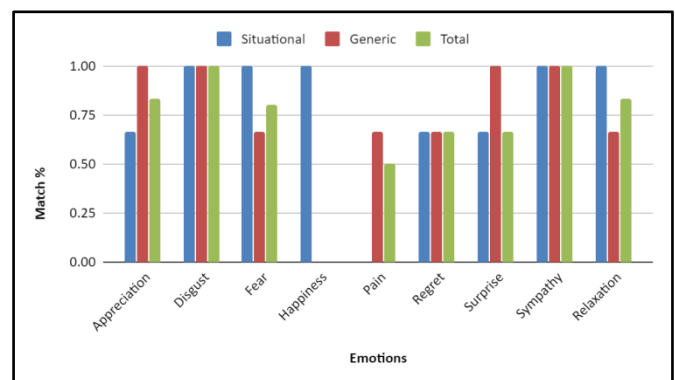
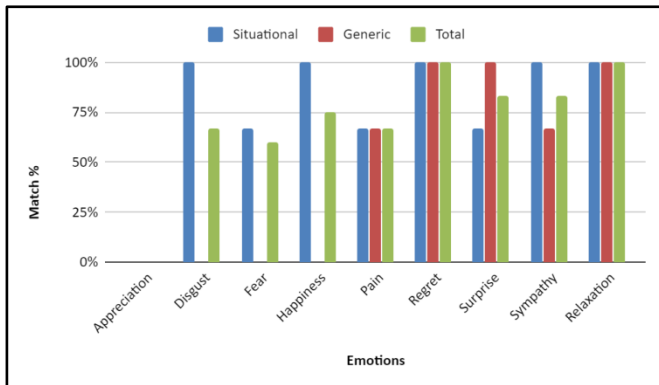


Figure 2(b): Comparison of matches in interjection usage across Bengali speakers

While both Telugu and Malayalam have 0% match in the generic category and above 50% matches in the situational category, Bengali has a 100% match in the situational category but none in the generic or across these two.





**Figure 2(c):** Comparison of matches in interjection usage across Malayalam speakers

- What catches immediate attention in the Malayalam data (see fig. 2(c)), is a complete lack of matches both within and across the response categories for the emotion 'appreciation'. Also note that, while there are no matches in the generic category for 3 emotions ('disgust', 'fear' and 'happiness'), there's a 100% match in 3 other emotions ('regret', 'surprise' and 'relaxation'). However, the matches across two categories of responses for majority of the emotions is around or above 75% in the Malayalam data, which is higher than the overall matches in both the Telugu and the Bengali data.
- When dealing with interjections in Indian languages, one has to keep in mind that most are primary interjections with no lexical content. These are non-words laden with specific intonation patterns that express the speaker's instantaneous mental/emotional state. This fact is reflected throughout the Indian data (see Appendix\_2), where the general lack of secondary interjection usage, both within or across categories for any of the emotions, is evident. For instance, responses for the emotions 'disgust', 'sympathy' and 'relaxation', across all the Indian languages included in this study, are essentially non-words with no particular meaning in any other context. Supporting this claim further, interjection usage across the Indian data shows a significantly lower frequency of 100% matches in the generic category (see figs. 2(a), 2(b) and 2(c)) when compared to the data from the Arabic-speaking countries (figs. 1(a), 1(b) and 1(c)).
- It is also interesting to note that the responses in the Indian data, display more instances of matches in interjection usage across the two categories: situational and generic than was the case for the Arabic-speaking countries. For example, data for the emotion 'fear' in Telugu, 'disgust' and 'sympathy' in Bengali and 'regret' and 'relaxation' in Malayalam shows a 100% match in interjection usage for both within and across the categories of responses.
- As seen in Appendix\_2, all the Indian languages use the same tokens interchangeably for various emotions. It could be due to the fact that there is a dearth of normative interjections, that is common to all speakers of the same language, to express every emotion. For instance, in the Bengali data, the participants used the words 'dhur', 'dhyat', 'jah', etc. for both situational and generic tokens for the emotion 'regret'. Similar instances can be seen in the Bengali data for 'happiness' and in the Malayalam data for 'sympathy'. Furthermore, interjections used in Telugu and Malayalam were

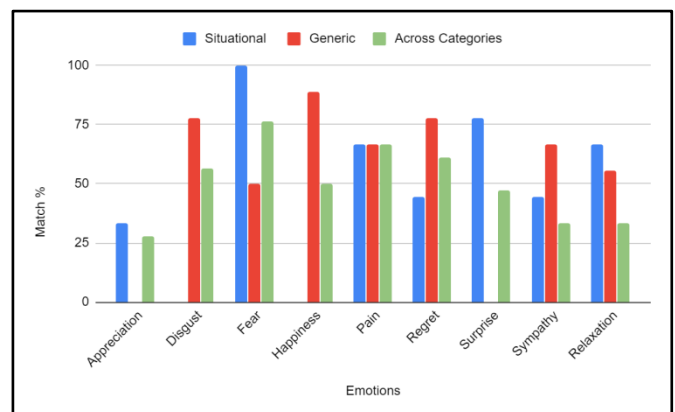
essentially the same words for a few emotions. For instance, in the case of both Telugu and Malayalam interjections, 'ayyo!' is used by participants for the emotion 'sympathy'. This token is a non-lexical item that is used as a primary interjection by speakers of both these languages but not by speakers of Bengali.

- It is interesting to note that for emotions like 'fear', 'happiness' and 'regret', the tokens were quite varied across all the languages from India that were included in this study. It could be due to the fact that the aforementioned emotions can be considered somewhat difficult to elicit natural reactions to when the stimulus is replicated in an artificial scenario.

### 5.3 Comparison of interjection usage across the Arabic-speaking countries vs. across the languages from India

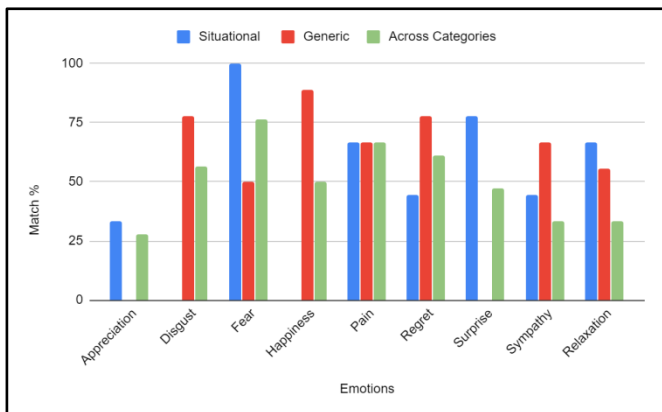
For this part of the study, interjection usage across the Arabic-speaking countries and across the languages from India was looked at. There were a total of 158 responses in the Arabic data; 4 less (participants did not submit a response) from the ideal 162: 18 per emotion, 9 per category and 18 across categories. Indian data had a total of 153 responses; 9 responses were not submitted by participants.

- The matches in interjection usage within both the categories seem consistently higher in the data across the Arabic-speaking countries (see Fig. 3(a)) as compared with the same in the data across the languages from India included in this study (see Fig. 3(b)).



**Figure 3(a):** Comparison of interjection usage across the Arabic-speaking countries (within and across categories)

- On the contrary, the Indian data displays greater percentages of matches in interjection usage across the two categories of responses than within those, in comparison with the Arabic data. This could be due to the fact that the Indian languages mostly use primary interjections to express emotions, which are often allomorphs of each other and are used interchangeably for different emotions and/or context. A more in-depth study of the semantic content of interjections in the Arabic data is required to draw more direct comparisons with the Indian data. Though that holds a possibility of leading to the stark contrast in the two sets of data, it is beyond the scope of the current study.

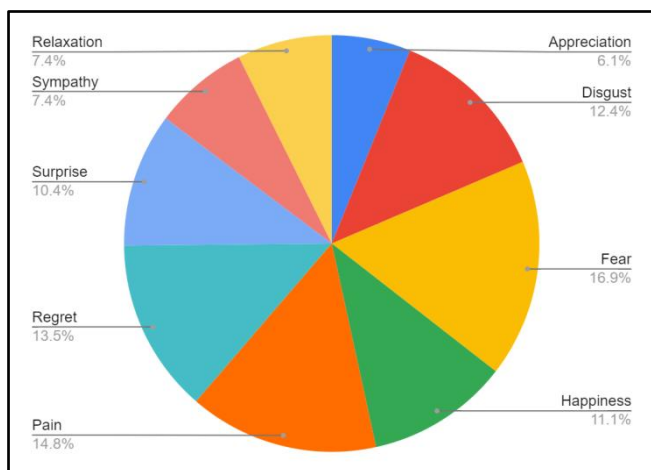


**Figure 3(b):** Comparison of interjection usage across the languages from India (both within and across categories)

- What is striking in the data representation of the matches in interjection usage across the languages from India (Fig.3(b)), is that there is a considerable lack in volume when compared to that of the Arabic data (Fig. 3(a)). In fact, none of the emotions shows a 100% match in either within or across categories in the Indian data, and almost all are below the 75% match cap.
- From an overall view, interjection usage seems to be more consistent across the countries speaking the same language than across different languages in the same country.

**5.4 Matches in interjection usage across the emotions**

For this part of the study, interjection usage across the emotions in the Arabic-speaking countries was compared with the same across the languages from India. There were a total of 158 responses in the Arabic data and 153 responses in the Indian data; 13 less (participants did not submit a response) from the ideal 162 per data set: 18 per emotion across categories.

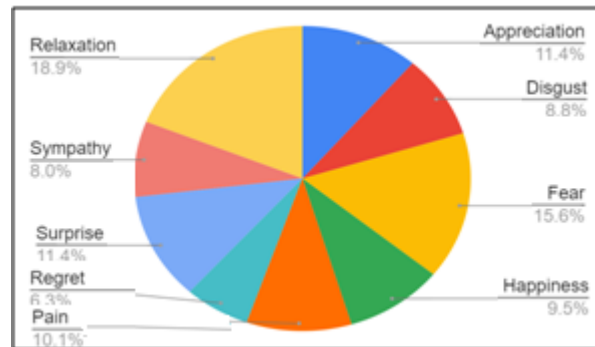


**Figure 4(a):** Matches in interjection usage across emotions in the Arabic-speaking countries

- In the Arabic data representation, the emotions ‘fear’ and ‘pain’ display the maximum percentage of matches in interjection usage across the two categories (see Fig. 4(a)). On the other hand, the emotions ‘appreciation’, ‘sympathy’ and ‘relaxation’ show the least percentage of matches in interjection usage. They each account for

lower than 10% of the total number of matches in interjection usage found across the data in Arabic-speaking countries.

- When compared to the Arabic data representation, the percentage of matches in interjection usage for the emotion ‘fear’ seems to be quite similar in the Indian data representation too (see Fig. 4b). In fact, the emotions ‘fear’ and ‘relaxation’ each display the highest percentages of matches found within an emotion in the Indian data. They each account for more than 15% of the total percentage of matches found in this data set.



**Figure 4(b):** Matches in interjection usage across the emotions in the languages from India

However, the majority of the emotions in the Indian data display less than 10% matches across categories, unlike the Arabic Data, wherein the majority of emotions display more than 10% matches across categories. This is suggestive of the fact that countries that were linguistically connected showed more matches in interjection usage, in almost all the emotions, than across the different languages that were spoken in the same country

**5.5 Terminology of Interjections**

- It may be interesting to note that many of the interjections across languages and countries seem to be an invocation to either ‘God’ (e.g.: ‘*ya Allah!*’ (Arabic), ‘*haayeBhogobaan!*’ (Bengali), ‘*Devuda!*’ (Telugu)); ‘father’ (e.g.: ‘*Baba go!*’ (Bengali), ‘*Abba!*’ (Telugu)) or ‘mother’ (e.g.: ‘*Amme!*’ (Malayalam), ‘*Ammo!*’ (Telugu), ‘*Yamma!*’ (Arabic)). These words seem to be rendered plurivocal by adding different intonations, across all the languages taken in this study, to give those the semantic shape they need.
- Additionally, some interjections across languages and countries seem to be allomorphs of the same morpheme (e.g.: ‘*ammo!*’ (Telugu), ‘*amme!*’ (Malayalam), ‘*o maa!*’ (Bengali), and ‘*yamma!*’ (Palestine), ‘*ya mamma!*’ (Lebanon).
- When dealing with interjections in Indian languages, one has to keep in mind that most are primary interjections with no lexical content. These are non-words laden with specific intonation patterns that express the speaker’s instantaneous mental/emotional state. This fact is reflected throughout the Indian data (see Appendix\_2), where the general lack of secondary interjection usage, both within or across categories for any of the emotions, is evident.
- It must be noted that all the Indian languages use the same tokens interchangeably for various emotions (e.g.:

'*ayyo!*' for both 'regret' and 'sympathy' (Telugu), and '*ayyo!*' for both 'pain' and 'sympathy'. It could be due to the fact that there is a dearth of normative interjections, that is common to all speakers of the same language, to express every emotion. Hence, speakers fall back on plurivocal interjections and assign various intonational markers to them to communicate a specific emotion. The addressee interprets the meaning of the interjection on the basis of the intonation used and the context it is used in.

## 6. Discussion

On the whole, it can be stated that the usage of interjections seems to be more linguistically influenced than by the other factors. In this study, though the Indian languages are all from the same geographical area, that is, the same country, the usage of interjections is quite varied across the languages. Telugu and Malayalam, however, show some similarities in their interjection usage, probably by virtue of (1) the states (where these two languages are spoken) being located geographically closer to one another than the state where Bengali is spoken, or because (2) Telugu and Malayalam both belong to the Dravidian group of languages, while Bengali is an Indo-Aryan language. Similarly, in the case of Arabic data, the geographically closer countries, Palestine and Lebanon, showed similar trends but differed quite a bit from Iraq. Therefore, it cannot be ruled out that geographical background also has its contribution in the pattern of interjection usage. But, it is evident in the data for this study that geographical proximity is still subsidiary to linguistic commonality in influencing interjection usage.

It must be noted here that owing to the availability of a limited data set, uniformity in participants in terms of age and gender could not be maintained. Since several researchers have reported speech token differences across age-groups and genders, there might be a possibility of there being such differences in interjection usage too. Also, whether interjection usage is more idiosyncratic in nature than influenced by either linguistic or geographic factors, could not be positively confirmed in this study. A larger data set could indicate the possibility of these phenomena, and a positive confirmation for the same would reflect that indeed, interjections are robust indexical markers of individual speakers.

Nonetheless, the current study suggests that both in the Indian and the Arabic context, interjections possess substantial discriminant power in a forensic phonetic scenario. They could prove to be dependable markers of linguistic, and hence, the cultural background of an individual in forensic speaker identification or speaker-profiling tasks. It is also of particular interest in the LADO

domain for asylum cases, as this study shows that even when all the three countries that were compared speak the same language, and all the languages that were analysed were spoken in the same country, analysis of the regionally and dialectically localised verbal tokens (that is, interjections) could indicate one's linguistic and hence, geographical background.

## References

- [1] I. Poggi, "The Language of Interjections," Esposito et al. (Eds.): Multimodal Signals, LNAI 5398, pp. 170-176, 2009.
- [2] E. Sapir, *Language*, New York: Harcourt, Brace, pp. 6, 1921.
- [3] A. Wierzbicka, *Cross-cultural Pragmatics: The Semantics of Human Interaction*, Mouton de Gruyter, Berlin, pp. 289, 2003.
- [4] C. F. Hockett, "The Origin of Speech", *Scientific American*, 203 (3), pp. 88-96, 1960. [Online]. Available: doi:10.1038/scientificamerican0960-88
- [5] F. Ameka, D. Wilkins, "Interjections," *Handbook of Pragmatics*, pp. 2-11, 2006.
- [6] Y. Jing, "English Interjections As A Word Class: A Tri-Stratal Description," *Indonesian Journal of Applied Linguistics*, 7 (1), 2017. [Online]. Available: <https://doi.org/10.17509/ijal.v7i1.6865>.
- [7] P. Kockelman, "The Meanings of Interjections in Q'eqchi' Maya: From Emotive Reaction to Social and Discursive Action," *Current Anthropology*, 44 (44), August - October, 2003.
- [8] T. Shannon, "Hallo! Ätsch! On the Semantic Analysis of Interjections," 10.13140/RG.2.2.30909.31209, 2018.
- [9] F. Y. Wang, "On Exclamatory Word- The Semantic Comparison of English and Chinese Interjections," *Journal of Peking University*, 1, pp. 17-20, 1999.
- [10] F. Ameka, "Interjections: The universal yet neglected part of speech," *Journal of Pragmatics*, 18 (2-3), pp. 101-118, 1992. [Online]. Available: doi:10.1016/0378-2166(92)90048-G
- [11] A. Wierzbicka, "The Semantics of Interjections," *Journal of Pragmatics*, 18, pp. 162-165, 1992.
- [12] A. Mao, "Conceptuality and Context-Sensitivity of Emotive Interjections," *Open Journal of Modern Linguistics*, 7, pp. 46, 2017. [Online] Available: <https://doi.org/10.4236/ojml.2017.71004>
- [13] P. Rose, *Forensic Speaker Identification*, Taylor and Francis, London, 2002.
- [14] Census of India, *Primary Census Abstracts*, Registrar General of India, Ministry of Home Affairs, Government of India, 2011. [Online] Available: <http://www.censusindia.gov>.



## Appendix\_1: Arabic Data

Country	Subject No.	Emotions								
		Appreciation	Disgust	Fear	Happiness	Pain	Regret	Surprise	Sympathy	Relaxation
Lebanon	1	<b>Bravo!</b>	<b>Buh buh!</b>	<b>Mahmi mamy!</b>	<b>Ribhet!</b>	<b>Ouf ouf!</b>	<b>Oh la non!</b>	<b>Ahlan wasalam!</b>	<b>Mahlesh!</b>	<b>Oh oh!</b>
		Allah-Allah	Agh agh!	Yiyiyiyi!	Yayaya!	Aye aye!	La la la la!	Hah!	Haram!	Khai!
	2	<b>Oh la la</b>	<b>Buh buh!</b>	<b>Ya mamma!</b>	<b>Oh Rbehet!</b>	<b>Ouf ouf!</b>	<b>Oh no no!</b>	<b>Ahla ahla!</b>	<b>Maa-lesh!</b>	<b>Uf ah!</b>
		Allah-Allah	Aakh iw!	Yiyiyiyi!	Yay yay yay!	Ah ah!	La la la la!	Hah!	Meskim!	Khai!
	3	<b>Wow!</b>	<b>Ouff!</b>	<b>Ya ammeh!</b>	<b>Hey Rbehet!</b>	<b>Akh akh!</b>	<b>Ya Allah!</b>	<b>Ahlan ahlan!</b>	<b>Ouf haram mehlesh!</b>	<b>Ah ah!</b>
		Allah-Allah	Ukh yuh	Ya delle yiyiyi!	Yay!	Aye!	La la!	Wow!	Haram!	Khay!
Iraq	1	<b>Wow!</b>	-	<b>Yama!</b>	<b>Wow!</b>	<b>Ach!</b>	<b>Ya illah!</b>	<b>Allahh!</b>	<b>Khatia!</b>	<b>Afesh!</b>
		Immm!	Eyaaa!	-	Ya Allah!	Ach ach!	Asif!	-	La la!	Oooh!
	2	<b>Ooh sphan Allah!</b>	<b>Ei ei</b>	<b>Yamma!</b>	<b>Oh Alhamad Alaha!</b>	<b>Ayi!</b>	<b>Ooff!</b>	<b>Ya Yallah!</b>	<b>Oooy!</b>	<b>Afisha!</b>
		Shakran!	Eyaaa!	Yamma!	Halooo!	Ohuy!	Haramat!	Ya Allah!	Ya ilahi!	Eyaa!
	3	<b>Wow!</b>	<b>Ta Hadikem</b>	<b>Yama of!</b>	<b>Eii!</b>	<b>Ach!</b>	<b>Yella!</b>	<b>Ya hella!</b>	<b>Khatia!</b>	<b>Afeesh!</b>
		Immm!	Ah ah	Of of!	Yahh!	Ach!	La-la!	Eii-eii!	La-la!	Eih-eih!
Palestine	1	<b>Fanan!</b>	<b>Yenaan shaklo!</b>	<b>Yama!</b>	<b>Yasalam!</b>	<b>Akh!</b>	<b>Akalna hawa!</b>	<b>Ahla wsahla!</b>	<b>Ya haram!</b>	<b>Oh moneesh!</b>
		Jameel!	Yaah!	Yama!	Yaay!	Ack!	Laa!	Wow!	Beean Allah!	Oh raha!
	2	<b>Fnaan!</b>	-	<b>Yama!</b>	<b>Yasalam!</b>	<b>Aaakh!</b>	<b>Akalna hawa!</b>	<b>Ana saeed feele sadeque</b>	<b>Ya haram!</b>	<b>Ooo mertah!</b>
		Jameel!	Yaeeaa!	Yama!	Yay!	Aah leh	Laa leesh!	Yaay!	Allah le!	Akheran!
	3	<b>Fanan!</b>	<b>Haylvan zay sahbou!</b>	<b>Yamma!</b>	<b>Yasalam!</b>	<b>Aaakh!</b>	<b>Akina hwa!</b>	<b>Meen ajani!</b>	<b>Ya haram!</b>	<b>Ooo mertah!</b>
		Jameel!	Yaeeaa!	Yamma!	Yayaya!	Aaakh!	Laa leesh!	Yaay!	Allah yena!	Akheeran!

Note: Situational category responses are in bold, while the generic category responses are in normal text.

## Appendix\_2: Indian Data

Language	Subject No.	Emotions								
		Appreciation	Disgust	Fear	Happiness	Pain	Regret	Surprise	Sympathy	Relaxation
Telugu	1	<b>Abba..bbah..</b>	<b>Ebbey! Chi!</b>	<b>Ammo!</b>	<b>Yay!</b>	<b>Aa!</b>	<b>Bah! / Chi!</b>	<b>Arrey!</b>	<b>Ayoo!</b>	<b>Abbaa!</b>
		Abba..bbah..	Chi! Beh!	Devuda! Ammo!	Yay!	Abba! / Aa!	Ayo!	Arey!	*pch*	Aah! Amma!
	2	<b>Bba..bba..</b>	<b>Chi! Thu!</b>	<b>Devuda! / Vammo!</b>	<b>Yaay!</b>	<b>Abba! / Devuda!</b>	<b>Cha!</b>	<b>Arey!</b>	<b>Ayyo..</b>	<b>Ahha! / Ammaya!</b>
		Ba ba	Beh...thu	Ammo! Devuda!	-	Abba! / Devuda	Cha cha!	Ah! Ammo!	*tch tch*	Aaha!
	3	<b>Wow! / Abboo!</b>	<b>Chi! Abbaa!</b>	<b>Ammo!</b>	<b>Wow!</b>	<b>Abba!</b>	<b>Ayyo!</b>	<b>Arrey!</b>	<b>Achocho! / Ayyo</b>	<b>Ammaa!</b>
		Wow!	Chii!	Ammo!	-	Abba!	Ayyo!	Oooo!	-	Aaha!
Bengali	1	<b>Baah!</b>	<b>Ishh!</b>	-	-	<b>Aah!</b>	<b>Dhur Baba</b>	<b>Hyaan?!</b>	<b>Aha re!</b>	<b>Aah!</b>
		Baah!	Ishh!	Ore Baba re!	Ki mowja!	Aah!	Bhul holo!	Ooo!	Ahaare!	Aah!
	2	<b>Ki darun!</b>	<b>Ishh!</b>	<b>Orey baba re</b>	<b>Aribbas!</b>	<b>O maa go!</b>	<b>Dhur Baba</b>	<b>Hyaan!</b>	<b>Aha re!</b>	<b>Aah! / Ki aaram!</b>
		Baah!	Ishh! / Chhi	Orey baba re	-	Aaah!	Ey Baba! / Jah!	Oh Maa!	Aha re!	Besh!
	3	<b>Baah!</b>	<b>Issh!</b>	<b>Oma go!</b>	<b>Arrey baas!</b>	<b>Oof!</b>	<b>Dhyat!</b>	<b>Orrey baba!</b>	<b>Aaha re!</b>	<b>Aaah!</b>
		Baah!	Eeshh!	Haaye Bhogobaan	Baah!	Maa go!	Eh baba!	Orrey baba!	Aaha re!	Aah!
Malayalam	1	<b>Albutham!</b>	<b>Ayye!</b>	<b>Aiyo!</b>	<b>Eeee!</b>	<b>Amme!</b>	<b>Shey!</b>	<b>Oh!</b>	<b>Aiyyo!</b>	<b>Aaha!</b>
		Albutham!	Ayye!	Amme!	Aha!	Aiyyo!	Sheyy!	Oh!	Paavam!	Ah!
	2	<b>Wow!</b>	<b>Ayye!</b>	<b>Uyyo!</b>	<b>Yay!</b>	<b>Ayyo!</b>	<b>Shey!</b>	<b>Hii!</b>	<b>Ayyo!</b>	<b>Haa...!</b>
		Waw!	Eww!	Uyyo!	Yay!	Aah!	Sho! / Ayyo!	Oh!	Ayyo!	Ah!
	3	<b>Ohh!</b>	<b>Ayee!</b>	<b>Ehh!</b>	-	<b>Ayyo!</b>	<b>Shey!</b>	<b>Neeyo!</b>	<b>Ashodaa!</b>	<b>Hahh!</b>
		Atheyo!	Chee!	-	-	Ayyo!	Shey!	Aano!	Achoda!	Hahh!

Note: Situational category responses are in bold, while the generic category responses are in normal text.