University of Hormozgan

Iranian Journal of Educational Research

Print ISSN: 1735 - 563X Online ISSN: 2980 - 874X

Homepage: http://ijer.hormozgan.ac.ir

Educational and Behavioral Research Center

Acoustic Representation of Prosodic Cues during the Production of Persian Irony and Sarcasm in Spontaneous Discourse

Raha Koochacki[1], Mohammad Hossein Sharafzadeh[2]✉, Ameneh Zare[3]

1. Ph.D. Student, Department of General Linguistics, Faculty of Humanities, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

2. Assistant Professor, Department of General Linguistics, Faculty of Humanities, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran, h-sharafzadeh@gmail.com

3. Assistant Professor, Department of General Linguistics, Faculty of Humanities, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran

Article Info

Article type:

Research Article

Article history:

Received 12 Dec. 2023

Received in revised form 25 Aug. 2024

Accepted 27 Sep. 2024

Published online 01 Dec. 2024

Keywords:

Production of Irony,
Sarcasm,
Pragmatic Strategies,
Prosodic Features,
Intonational Contours,
Spontaneous Discourse

ABSTRACT

Objective: The present study has tackled to assess the interaction between lexical-syntactic and prosodic cues of irony, and the fact that how the cues are represented during the production of irony and sarcasm (as one of highly functional irony subtypes) in spontaneous discourse through an auditory modality.

Methods: For this purpose, spontaneously produced ironic and non-ironic (baseline) utterances via pairs of friends in conversational dyads during and after watching video stimuli were elicited and analyzed for pragmatic, prosodic contrasts. Thereupon any pragmatic strategies (irony subtypes), the observed lexical-syntactic cues accompanied by the prosodic features of the target utterances coded considering TOBI system and using Praat software, were annotated manually in different ELAN tiers.

Results: The statistic results showed higher values of auditory characteristics in ironic utterances versus baseline utterances. The ultimate probing indicated that the prosodic cues, in production task of Persian irony, operate together as a single system and contribute to intelligibility and detection of an ironic speech even in the absence of verbal markers. Furthermore, there were no correlated patterns between certain types of irony and certain pitch patterns and intonational contours.

Conclusions: The results of the present research can be used in Persian language classes in schools and universities.

Cite this article: Koochacki, R., Sharafzadeh, M. H. & Zare, A. (2024). Acoustic representation of prosodic cues during the production of Persian irony and sarcasm in spontaneous discourse. *Iranian Journal of Educational Research*, 3 (4), 299-322.

DOI: https://doi.org/10.22034/3.4.299

© The Author(s).

Publisher: University of Hormozgan.

DOI: https://doi.org/10.22034/3.4.299

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

Introduction

In our culture irony and sarcasm are the most common non-literal communicative forms. Irony can convey positive or negative meaning. However, sarcasm is considered as an irony subtype that has a negative, critical and hostile meaning. Sarcasm is an attitude with the content that "I don't mean this" and in which a speaker deliberately teases another person or a point of view that indeed "does mean this" or shows his hatred towards it (Haverkate, 1990; Gibbs, ٢٠٠٠). People like Capelli and et al. (1990) use the terms sarcasm and irony in a very similar sense. Nevertheless Haiman (1998) dissociates these two terms by stating that sarcasm must be intentional and transmitted by people, while irony may address situations and can be unintentional. "Attitude" is a key term in the definition of sarcasm and must be distinguished from the word "emotion". "Attitude" is transmitted in an indirect way through cues in speech, hence there is a need for a linguistic investigation of textual-contextual information, intonation, and pragmatics and semantics, whereas "emotion" is interpreted directly through speech signals (Wichmann, 2000). Furthermore, from Leach's politeness perspective, sarcasm is clearly a form of impolite speech that is intended to be conveyed as an insulting behavior through an utterance that would naturally be polite or sincere. Therefore, sarcasm is a mock politeness or a meta-strategy that uses politeness to communicate impoliteness (Leech, 1983; Culpeper, 2005).

As matter of fact, Irony and sarcasm cues can be lexically, such as the use of exaggerated adjectives and adverbs, and also nonverbally, such as facial expressions and through Prosodic modulations marked. . One of the basic cues of sarcasm is the "sarcastic tone of voice" which constitutes it an intonational misfit where the prosodic features are affected by the negative spin that the speaker intends to put on his message to distinguish it from what is literally positive and sincere (Cruttenden, 1984). The sarcastic tone of voice is placed in the attitudinal intonation subgroup, which conducts us to the speaker's behavior (i.e. intentions and perceptions) in a context (Wichmann, 2000).

Haiman (1998) asserts that both segmental and suprasegmental features change when the speaker tends to be more sarcastic than sincere. These signals, which can be different across languages, accents, and situational varieties, may be combined in time with other features such as semantic cues at the word or phrase level, consist of durational modifications; Nasalization of sounds and

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

flattening or exaggerating ʃ0 rises and falls. Context cues that are non-verbal in nature, such as eye movement and facial expressions, can also be changed in sarcasm (Rockwell, 2000 a, 2005).

Literature of Review

According to specific production studies in the field of prosody of sarcasm, Fonagy's (1971) studies on Hungarian language indicate a drastic increase in stressed syllable duration, intensity, and fundamental frequency range (ʃ0 range) in a sarcastic speech. Such durational increases, which lead to a decrease in speech rate, are relatively constant across completely unrelated languages (e.g., Adachi, 1996 for Japanese; Haiman, 1998 for English, Greek and Tagalog; Culpeper, 1996 for British English; Cheang & Pell, 2008 and 2009 for Canadian and Cantonese).

In addition, a special type of sarcastic intonation, which is "insulting" in nature, can be observed when the fundamental frequency (ʃ0) in the stressed syllable of certain words that are the key to conveying sarcasm is relatively "low" (Winner, 1988). This idea of a low-flat ʃ0 movement in the stressed syllable is typically found in sarcasm and is intended to moderate the enthusiasm that positive words normally convey. However, on the contrary, sarcastic messages containing increased ʃ0 excursions were observed in Attardo et al. (2003) and Laval & Bert-Erboul (2005). Such complexities are further found in Haiman (1998), whose English data depict a high-low ʃ0 movement accompanied by increased intensity in echo utterances of sarcasm. He acknowledges similar activity of ʃ0 in sarcastic messages, where the duration and intensity may indicate a sincere and literal interpretation. Moreover, Ladd's (1978) indicate that exaggerating in ʃ0 may be employed to put a sarcastic twist on a predictable message.

From another perspective, Rockwell (2007) compared acoustic and perceptual approaches for sarcasm. He found that prosodic cues in an acoustic modality discriminated sarcastic utterances slightly more strongly than non-sarcastic literal utterances (for exemplary perception-based investigations, see Bryant and Fox Tree (2002, 2005), Rockwell (2000b, 2007), and Voyer and Techentin (2010). Based on auditory analysis, he observed that ʃ0 movement and its range differed significantly between sarcastic and non-sarcastic conditions.

At last, Bryant (2010) is one of the first researchers to explain the prosodic contrast between neighboring phrase units through a detailed auditory analysis. He claims that the issue is a crucial step in the field since such contrasts are vital to a better perception of how undertones interact with speaker intent in speech.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

In general, there are mixed findings about the prosody of sarcasm in studies to date, which are effectively summarized by Cheang and Pell (2009:1394): A comparison of various languages confirms that many of them employ prosodic elements to convey meanings that are not to be interpreted literally and he way in which such prosodic parameters are implemented seems to vary between languages. In their research, they focused on the importance of prosody features such as intonation, utterance rate/rhythm, phrasing, etc. as indicators or cues for detecting sarcasm (Cheang & Pell, 2008).

Considering real audio cues, it seems that sarcasm can be encoded in utterance through general various manipulations in acoustic parameters such as fundamental frequency (ʃ0), amplitude, speech rate, voice quality, vowel hyper-articulation (Attardo et al., 2003; Bryant and Fox Tree, 2005; Cheang and Pell, 2008; Rockwell, 2000; Scharrer and Christman, 2011). On the other hand, an old study on the prosodic features of sarcastic speech in a corpus of indicative statements in French (Ben Jannet, 2012; Loevenbruck et al., 2013) demonstrates that sarcastic productions are discriminated by lengthening the utterance, instead of increased modulations of ʃ0 and an overall rise in the pitch level and span.

Overall, the findings obtained from various studies regarding prosodic correlates of irony and sarcasm with pragmatic properties lack consistency. For instance, whereas some reports indicated reductions in standard deviation (SD) of ʃ0 for ironic and sarcastic utterances against sincerely utterances (Attardo et al., 2003; Cheang and Pell, 2008; Fongy, 1971, 1976), the others demonstrated an increase in standard deviation of ʃ0 in ironic and sarcastic utterances (Haimann, 1998; Cutler, 1974, 1976). As a result, the void of a systematic and full-scale study by determining acoustic/prosodic signals is thoroughly felt and on the other hand, considering that most of the studies conducted in this scope are on the English-speaking context, the study of this phenomenon in other languages is necessary to explain more about the Cross-cultural similarities and differences is proper. In addition, in order to explain all available methodological and contextual constraints, a systematic study of this phenomenon in spontaneous discourse appears essential.

Therefore, the present study, in particular, investigates the acoustic/ prosodic markers of irony and sarcasm under the conditions of spontaneous discourse in Persian language. The main aim of the researcher is to clarify whether irony in general and sarcasm in particular are marked acoustically in Persian language? And particularly, this study tackles to examine and explain prosodic

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

correlates of irony and sarcasm, such as levels and variability of ʃ0, duration, intensity, etc. …in detected and encoded instances during the production of irony and sarcasm in spontaneous discourse and that how does the acoustic measures such as fundamental frequency, duration, intensity, etc. based on changes between sarcastic and sincere attitudes change in Persian native speakers in spontaneous discourse?

Material and Methods

3. Test Method

3.1. Participants

A total of 20 educated Persian speakers (18 men and 2 women, with a mean age of 20.35 and a standard deviation of 2.724) from Fars province region (mainly students in Firuzabad universities) participated in this study. All volunteers in the study in pairs (10 pairs in total) and according to the investigations carried out in previous studies (for example, Gibbs, 2000), with precondition having a family history and friendship were selected. All participants were native speakers and their dominant language was Farsi. Only three speakers also spoke Turkish. According to the participants' self-reports about the amount of time they spoke in Farsi every day, the dominant value was 89.98% (stdev 16.673). The participants declared their unconditional consent to audio and video recording and permission was obtained from all of them to apply their data and information for research and educational purposes. Each of the participants was paid a decent wage for participating in the experiment.

3.2. Materials

The stimulus materials of the experiment consist of 1) two video chains (labeled as video (a) and video (b)) (to see two motionless images from each video, refer to Figure 1) and 2) a set of 8 sentences related to the videos (4 sentences for each video), included in 2 distinct cards (see examples 1 and 2 in the section below). The aim of selecting video chains and sentences is to provoke the incongruent contextual settings and to extract the participants' reactions after watching each video. With the prediction that these types of sentences will lead to natural sarcastic responses and induce to produce sarcastic statements (in the case of video A) and non-sarcastic statements (in the case of video b). Therefore, while the contents of the video and sentences for

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

video (a) were incongruous and inconsistent, in video (b), they were rationally congruent and consistent.

Example 1: This is called a subtle acting!

Example 2: He's a bright future!

With respect to Curco (2000) and Morreall (1989) regarding the close relationship between the cognitive processes involved in producing and detecting humor and irony, two videos related to the inconsistency and incongruity of the same situation (here the sitcom Dracula and the video clip Nuisance) were selected. At first, video (a) (4́: 30̋), a group of amateur actors with an unprofessional performance; And video (b) (5́:22̋) depicted a group of professional actors with a proficient performance technique. See the two image frames in Figure 1).



Figure1. Motionless images from video A (left frame) and video B (right frame)

3.3. Procedure

Audio and video recording was done in a quiet room at the test location in the counseling center of Azad University of Firuzabad. The participants were enrolled for this experiment in pairs and knowing this precondition that they should have a relationship of friendship or family ties with the other person. After the participants entered the test site, they were randomly selected as "Speaker A" and "Speaker B". As can be seen in Figure 2, the two participants sat on the designated chairs facing each other with a distance of about 1.5 meters. There was a laptop computer equipped with earphones in front of each participant, next to the computer there was a card containing 4 prompt sentences (participant "A" had the four Video A sentences and Speaker B the four Video B sentences). Video cameras (a Camcorder Hc-X1000, a Sony HD-*Handycam* HDR-AX2000 and a Nikon AF-PDX NIKKOR camera) were installed in the test location, the focus of two of the

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

cameras was aimed at two speakers and the third was recording a long shot of the entire scene. It is noteworthy that the purpose of using cameras and recording images in the test site was to collect the visual data in order to investigate gestural signals in the production and perception of irony and sarcasm where in another study, the researcher dealt with it in detail. Also, this experiment was carried out using two AKG A-27D Delegate desktop cameras installed next to laptops and with the help of a hidden digital voice recorder called Lander LD75-i installed in behind the table, recorded the voice. The participants were unaware of the real purpose of the study and they were told that the aim of this experiment was to examine issues related to public communication. To make the conversational interaction as natural as possible, no instructions about seating height, body posture or gestures were given. The video stimuli were presented in the same order and alternately for all ten pairs of participants, and each was given the written instructions below:

There are 2 video files on your laptop desktop. Watch them simultaneously and in order, and discuss about what you have seen. Your task is not to describe the content of the videos, but to evaluate what you have seen, commenting freely, criticizing, praising or even joking. You will listen to the audio track using only one earphone, so you can hear what your partner says and share impressions with him/her. When you finish watching Video A, close the lid of the laptop and do two things. At first, exchange the general impressions about this video. Then the participant who has the card corresponding to video A must read aloud the set of sentences on that card: as each sentence is read out, you must both react to it and make comments When you have finished, repeat this procedure with Video B.

Then the participants were left alone in the experiment room and they were asked to call the experimenter after completing the experiment. All conversations were audio-visually recorded by three cameras and the voice recorder.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]



Figure 2. A medium close shot of the location and placement of the participants and test equipment to record sound and video under spontaneous conditions.

3.3.1. data coding

First, sarcastic utterances were extracted from 10 conversations (obtained from spontaneous exchanges and reactions to the recording of sarcastic prompt sentences). Whenever possible, any utterances that immediately preceded the target sarcastic utterances (henceforth, baseline utterances) were also extracted. The selection of sarcastic statements was made in accordance with the following wide and proposed definition of Gibbs (2000: 13): "Each form of irony minimally reflected the idea of a speaker providing some contrast between expectation and reality". It should be noted that it is possible for speakers to start producing the sarcastic cues in utterances prior to a sarcastic utterance. Therefore, it was decided to examine the sentences immediately before the target sarcastic sentence. This issue was due to 2 main reasons: (1) The researcher follows Bryant's (2010) method to compare sarcastic and non-sarcastic sentences, so the results obtained are directly comparable with her studies and (2) if the researcher finds a clear difference in sarcastic cues between the target sarcastic utterances and the utterances immediately before them (which may be some show weaker signals), the results from special correlates of target sarcastic sentences will be stronger. Then the baseline utterances and the target sarcastic utterances were transcribed orthographically and a number of lexical-syntactic and auditory cues were manually annotated using the ELAN program (version 6100) (lausberg and Sloetjes, 2009) by the researcher. All the pragmatic strategies (irony subtypes) and lexical-syntactic cues and visual signs observed were annotated in different tiers of the ELAN program (see Figure 3). Next, the prosodic features of the target utterances (in the form of a wavegram) were encoded using the program praat (version 4/5) (Boersma and Weenink, 2008) and then automatically imported into ELAN program.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

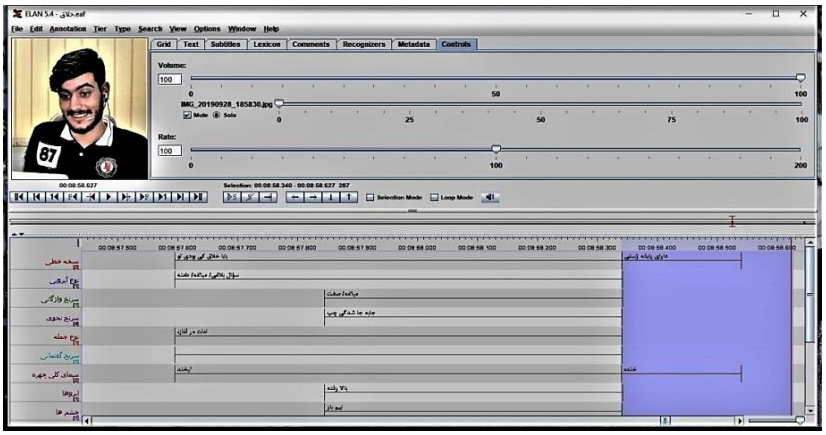


Figure 3. Part I. An example of lexical-syntactic and discourse annotation with target sarcastic sentence "Dear me! Who were you creative for!"

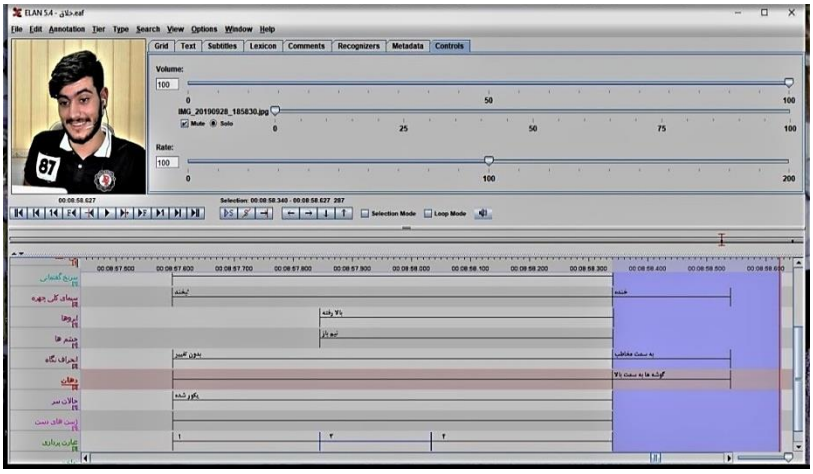


Figure 3. Part II. An example of visual and prosodic annotation with target sarcastic sentence "Dear me! Who were you creative for!"

In the following section, a brief explanation of the applied coding in the study is presented for any tier included in the ELAN program.

Tier 1. Orthographical transcription. The first tier consisted of performing an orthographically transcribed phoneme of the target utterances.

Tier 2. Pragmatic coding.

Following Gibbs' proposal (2000), five irony subtypes were used in this study:

1- "Sarcasm", where speakers convey a negative meaning with positive words. In this study, this type of irony was observed as one of the most prototypical irony subtypes. For instance, in a part of this experiment, one of the participants watching the video (A) addressed the main role actor,

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

who had a very artificial and weak performance, with the term "Master!" and elsewhere, in the same situation, used the term "creative!"; 2- Hyperbole, where the speakers stated non-literal meaning (i.e. figurative meaning) to exaggerate the reality of that situation in order to have a greater effect on the listener). In this study, a participant used the phrase "I burst out laughing" in an ironic form, which definitely meant nothing but to express the intensity of laughter; 3- "Understatement", where speakers conveyed the ironic messages with a much less valuable expression than their intended appearance. For example: "You're still too young to understand this!"; 4- "Jocularity", where the ironic meaning was conveyed with teasing and humor. For example, in this study, one can refer to the phrase "Oh my God! Cutie!"; and 5- "Rhetorical questions", where speakers ask questions not to elicit an answer but to emphasize an obvious point that conveys a critical or humorous implication. For example, in folk literature: "They must hunger in winter that will not work in summer" and as an example in classic literature: "Those who turn lead into gold...will they ever our sight behold?" And in this corpus, we can mention examples like "Does he know that he's a genius?" or "Do you think I should audition for him?"

Tier 3-6: lexical-syntactic coding

Tier 3. This tier were used for annotating the presence of exaggerated adjectives, words and expressions such as "genius!", "creative!", "I love him", "very" as well as mitigation[1] words and expressions such as "a little", "perhaps" , maybe, " Do you think…?", "Do you feel…?", "I don't believe", "I think" (refer to scharrer and et al., 2011).

Tier 4. This tier was used to annotate the presence of superlative or diminutive prefixes, suffixes and adverbs. Among the examples of statements used in this study are: "very" or the suffix "-est" in the word "the greatest" or the prefix "im-", respectively, in sarcastic utterances "In her role she's very immersed"; "He's the greatest discovery of the century" and "What an impeccable gesture!"

Tier 5. This tier was used for annotating Right/Left Topicalizations. Here it can be referred to the sarcastic utterances such as "So talented, this director!"; or " Dear me! You were creative for who"!; or " In her role she's very immersed " (for more details refer to EScandell-Vidal et al., 2014).

Tier 6. This tier was used to annotate types of commonly used sentences in ironic utterances included in the study. After the investigations, 4 types of sentences were considered as the most used in this experiment: 1) Sentences with discourse markers from the type of particle at the

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

beginning of the sentence 1) Sentences with discourse markers from the type of particle at the beginning of the sentence; 2) Exclamation sentences with the structure of the word "what" at the beginning of the sentence and "exclamation mark" at the end of the sentence; 3) declarative sentences; 4) Sentences with the content of tag questions.

Tier 7: Discourse coding

This tier was used for annotating 1) *code-switching and code-mixing*[3] *features*. As an example in this study: "This is called a superstar!", Where the combination of Persian and English is used in a phrase for a specific effect; 2) *Direct speech in a language other than Persian*. For example in this study: "The poet says: Bu Hara O Hara!" (with an English equivalent of "This is a far cry from that"), where the format is Persian, while the direct speech is in Turkish; and also 3) *discourse markers*[2] such as "Well!"; "Of course!"; "Right!" (See Ruiz Grillo; 2008; Munoa-Barredo, 1997; Schiffrin, 1987).

---

[1] *Mitigation* : is "a reduction in how unpleasant, serious, etc something is". Mitigated speech as a linguistic term describing deferential or indirect speech inherent in communication between individuals of perceived high power distance has been used for at least two decades. The term has recently been popularized by Malcolm Gladwell defining mitigated speech as "any attempt to downplay or sugarcoat the meaning of what is being said".

[2] *discourse markers:* Using words or phrases that have no relation to the original and lexical meaning of that word or phrase. In other words, a word that is devoid of any lexical meaning. Among the criteria for recognizing the role of discourse views, we can mention establishing a connection with the previous content without entering into the syntactic structure and propositional meaning.

[3] *code-switching and code-mixing*: Alternating between two or more languages or language varieties/dialects in the context of a single conversation. Using elements of more than one language when conversing in a manner that is consistent with the syntax, morphology, and phonology of each language or dialect. Truth be told, many people use the terms Code-Switching and Code-Mixing interchangeably. Some linguists, however, make a distinction in which Code Mixing refers to the hybridization of two languages (e.g. parkear, which uses an English root word and Spanish morphology) and Code-Switching refers to the movement from one language to another.

Tier 15-21: Prosodic coding

Tier 15. Phrasing.

Following the Cat-TOBI hypothesis (Prieto, 2014)[1], this feature was annotated in the pause subtypes (that is, the boundary strength level of prosodic groups) were annotated: the prosodic elements containing clitics with word content labeled "0"; word chains labeled "1-2"; The word end or intermediate phrases labeled "3" and the end of the rhythmic phrases (rhythmic phrase (word groups) end) labeled "4".

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

Tier 16. Tonal nuclear configurations.

Once again, following the Cat-TOBI hypothesis (Prieto, 2014), boundary tones (those tones related to intonation boundaries) and pitch accents (those stresses that correspond to stressed syllables) were labeled.

Tier 17. Voice quality.

In this tier, the perceptual features of voice quality (labeled 'Creaky', 'Falsetto' or 'Breathy') were annotated and verified by exploring their acoustic correlates using Praat program (Boersma and Weenik, 2008).

And ultimately, following Bryant (2010), below values from both situations, i.e. target sarcastic utterances and baseline utterances), were extracted and in tiers 18, 19, 20 and 21, respectively, for the average pitch (mean ʃ0) on the Hertz scale ; pitch variability (or standard deviation values) on Hertz scale; average loudness (voice intensity) on decibel scale (dBs) and mean syllable duration, i.e., speech rate (equivalent to MSD or mean SD) on ms scale ere annotated. In order to modify variability in ʃ0 measures among speakers, ʃ0 measures were turned into semitones (against 1 HZ). mean syllable duration (MSD) was considered as a measure of speech rate and was computed through dividing the duration of the entire target utterance of the entire target utterance (on ms scale) by the number of syllables.

3.3.2. inter-rater reliability

In order to assess the reliability of (a) detecting sarcastic utterances and (b) pragmatic and prosodic coding of target sarcastic utterances, as explained above, an inter-rater reliability test was conducted with a subset of 20% of the data. Three independent raters (including the researcher, a psychologist and a linguist) labeled a random selection of data after the instructions explained in the previous section. Since the total duration of the recordings was up to 3 hours and 55 minutes (a total of 235 minutes), the reliability test involved 47 minutes of video (20% of the total play time) was taken into account. For pragmatic coding as well acoustic coding, a random selection of 16 target sarcastic and baseline utterances, which once again composed of the total sum of 20% of the data, was carried out.

At this stage, it was used the statistical measurement Fleiss fixed marginal Kappa (Iverson and Goldin-Meadow, 2005) due to the computation of the agreement measure of the raters in the classification probably predicted during the research and also because of the presence of 3 raters

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

in this study. The statistical instructions of Fleiss (1981: 214) characterize the kappa coefficient over 0.75 as excellent, 0.40 to 0.75 as fair to good and below 0.40 as poor. The Fliss statistic obtained for the detection and classification of ironic utterances was 0.71 and 0.72 respectively; for literal cues (considered overall), it was 0.89; for prosodic cues, it was 0.48 in tone nuclear configurations, 0.85 in phrasing and 0.87 in voice quality. It is thought that the above scores indicate a principled agreement between the raters and therefore validate the markings made in this corpus.

4. Data analysis

Among a total of 80 utterances obtained from spontaneous conversations between 20 pairs of participants in this study, a total of 54 sarcastic utterances were extracted from the database. Of these, 48 target sarcastic utterances from their previous utterances (i.e. without overlapping items) were available for analysis. In this section, it is dealt with the results

[1]*The Cat_ToBI proposal* : consists on a description of the prosodic and intonational structure of Catalan within the Autosegmental-Metrical (AM)
framework (Pierrehumbert, 1980; Gussenhoven, 2004, among others). This proposal includes an analysis of the phonetic realizations and distributional properties of the phonological intonational patterns found in Catalan, as well as the description of the intonational realization of different pragmatic meanings.

from analyzing semantic, lexical-syntactic and acoustic data and an exhaustive report of all investigated variables that were obtained in order to characterize the corpus as well as comparing the results of this study with those obtained from the reports previously mentioned in literature review.

4.1. Irony Subtypes

In Figure 4, the frequency of occurrence (%) of 5 irony subtypes found (based on Gibbs categorization) (2000) in the 48 ironic utterances was depicted. The most common and usable irony subtypes was jocularity with 31%, followed by a small margin, sarcasm with 27% and hyperbole with the occurrence of 24%. Rhetorical questions with 11% and understatement with 5% were in the next ranks in terms of frequency of occurrence.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

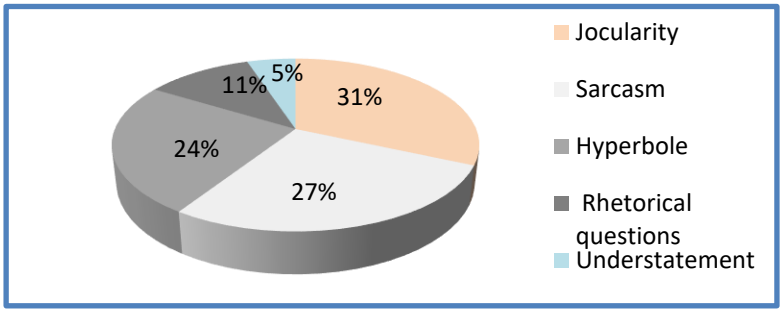


Chart 1. Frequency of occurrence of irony subtypes found in the 48 target ironic utterances under the conditions of spontaneous discourse.

4.2. Lexical-syntactic cues

In this section, as well as in the following sections, the results will be presented through a comparison of the target ironic utterances with baseline utterances (i.e., utterances that were uttered immediately before the target ironic utterances). As predicted, lexical-syntactic cues, or discourse verbal irony, appeared more frequently in target ironic utterances. A series of chi-square tests indicated that the presence of lexical signs in ironic utterances was significantly different compared to bas utterances ($x2(1) = 6/04\, p < 0/05$). We also achieved interesting results in lexical-syntactic studies. In this study, speakers, in order to produce irony, used a variety of special sentences to convey ironic intonation in a spontaneous situation (see diagram 1 in the section below). In general, in this study, there were 4 types of sentences with special lexical and syntactic combinations in ironic production:

1- Sentences with discourse markers from the type of particle at the beginning of the sentence. For instance, in this study: "Divine! Grant us more of these geniuses!" and or "Dear me! You were creative, for who!"

2- Exclamation sentences with the structure of the word "what" at the beginning of the sentence and "exclamation mark" at the end of the sentence. For example, "What a piercing glance!"

3- Declarative sentences. These types of sentences did not have any distinctive features except that at the time of ironic production, they consisted of words that contained cues of sarcasm. An example of a participant's response to video stimulus (A) is: "Definitely worth watching once!"

4- Ultimately, Sentences with the content of tag questions. The fourth type of the participants' responses was in ironic conditions. For example, in this corpus: "Not at all" like him, right? "

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

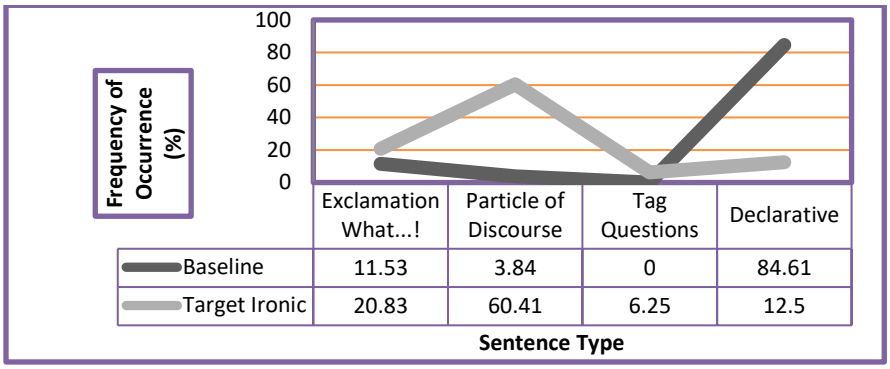


Chart 2. distribution of Occurrence frequency (%) of 4 commonly used sentence types in baseline and target ironic utterances.

Comparing these percentages, it was found that the sentences containing discourse particle (at the beginning of the sentence) had the highest use (60.41%) in the production of ironic utterances among the participants of this study, whereas in the production of baseline utterances, as expected, the highest rate of occurrence (84.61%) was assigned to declarative sentences. According to Kreuz and Caucci (2012), on the one hand, the production of sentences containing discourse particle has a lower degree of difficulty than other types of sentences, and on the other hand, discourse particle signs can have such sarcastic implications that these words themselves may be considered as cues for an ironic interpretation. Among the other three types of sentences, the second rank of occurrence in both ironic and baseline utterances was assigned to exclamation sentences with the sentence structure "What...!". No possible explanation has been provided in the literature review explored in the process of this study. A hypothesized explanation for this could be that the construction "what.....!" is commonly "used as a sarcastic structure in contemporary film and theater. However, there is no empirical data to support this claim. Tag questions in both ironic and baseline utterances received the last rank of applicability. Despite the statements of Kreuz and Caucci (2012), who put this type of sentences in the category of sentences with a low degree of difficulty for producing irony and sarcasm (in other words, producing responses with signs of discourse particle and tag questions than other sentences for producing sarcasm will be easier) nevertheless, in this study, the participants more preferred sentences with discourse particle and exclamation sentences with the structure "What....!" to produce an sarcastic intonation. This issue can be justified by considering that the sarcastic intonation may be different in each country, region, language and probably for each person.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

4.3. Discourse cues

Regarding discourse cues, although a similar percentage of ironic and baseline utterances had discourse cues, this is due to the fact that these utterances in both conditions had used a wide set of discourse markers. However, when specific types of discourse cues were analyzed, out of 48 available ironic utterances, 7 utterances (14.5%) applied code-switching or code-mixing and only 1 ironic utterance (2%) used direct speech to another language other than Persian (in this study, Turkish language, that is, the expression: "The poet says: "Bu Hara O Hara"). On the contrary, none of the baseline utterances used code-switching or code-mixing, as well as direct speech in another language other than Persian.

4.4. prosodic cues

In the TOBI system (the system of Tones and Break Indices), a standard system in the labeling of prosodic features based on the theory of auto segmental-metrical (AM) phonology, an utterance is intersected and labeled in 4 different tiers. These tiers are: 1) Orthographic Tier; 2) Tone Tier; 3) Break index tier and 4) miscellaneous tier. Of these four tiers, the orthographic tier is the tier where the transcribed form of utterance is observed. The tone tier and break index tier contain linguistic prosodic markers, which are discussed concisely in the following section. The miscellaneous tier is a tier used to display nonverbal sounds such as sneeze, laugh, cough, etc. or to add the descriptions.

4.4.1. Tone Tier

Considering the TOBI system, high tones (H) and low tones (L) in the Persian phonetic system are tantamount to two phonemes which appear alone or compound in pitch accents and edge tones. According to Eslami (1384: 21-22), the intonational elements in Persian include 4 pitch accents, 2 phrase tones and 2 boundary tones, which each of them gives a particular intonational and contextual meaning to the speech. Indeed, these intonational elements are the morphemes that appear in the form of tones H and L and convey their own special meaning to utterance. Pitch accents through changing in the level, direction and shape of the intonational contour in the lexical stressed syllable of words undertake the function of types of foregrounding and lateral tones, unlike pitch accents, do not contribute to foregrounding in speech; Rather, they play the role of boundary element in the intonational phrase. Pitch accents emerge either in the form of one tone (H* or L*) or in the form of two tones (L* +H or L+ H *), but the lateral tones manifest only in the form of

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

one tone (for phrasal tone (H−and L−) and for boundary tone (L% and H%)). (Here, "*" stands for the place of pitch accent, "−" for phrasal tone, and "%" for the border of intonational phrase).

4.4.1.1. Tonal nuclear configurations

As expected, the typical tonal configuration of a broad focus statement (for example L^* L%) was more frequently found in baseline utterances than in ironic target utterances (89% in baseline utterances and 63% in ironic target utterances). By contrast, ironic utterances were produced with more prominent configuration of emphatic and pragmatic meanings. (for example, $L+H^*L\%$, $L^*HL\%, L^*H\%$ یا $L+H^*LH\%$). In addition, ironic utterances were produced with interrogative nuclear configurations as in the case of ($L^*H\%$ and $L+H^*H\%$. The researcher did not observe any correlation between the nuclear configuration type and the irony subtype of the utterance.

4.4.2. Break index Tier

The values in this tier indicate the prosodic cohesion degree of a word with its next word in the utterance. The numbers from 0 to 4 respectively represent the coherence degree of the words in speech. The number 0 stands for the border of merging two words into each other. The number 1 represents the normal break between the words. The number 2 stands for the abnormal break of two words. The number 3 represents the border of "intermediate phrase*" and the number 4 stands for the border of "intonational phrase" (Eslami, 1384: 19).

4.4.3. Phrasing

The results from this study indicate that ironic utterances contained higher rates of prosodic breaks (for instance, those ironic utterances with a '3' or a '4' break index value) than baseline utterances (53% in ironic utterances vs. 16% in baseline utterances). A chi-square test demonstrated that the difference between the two groups was statistically significant ($x^2(1) = 5/90,\ P < /05$).

4.4.4. Pitch, intensity and duration measurements

In order to audio analysis, Praat software (version 6100) was employed. For this purpose, after opening the intended audio file and selecting the intended episode of audio signal, the parts before and after the selected signal that were not the target of the analysis were removed from the audio file. Thereafter the analysis of audio data was conducted on the rest of parts. On average, 2 minutes of each audio file were analyzed, which contains ironic target utterances and baseline utterances.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

It is worth noting that in order to analyze speech of the subjects while talking with their teammate, after opening an audio file in Praat, the speech sample was precisely examined and their teammate's voice was isolated from that conversation to be used in subsequent analyses. Figures 5 and 6 are examples of audio analyses based upon TOBI system which examine the prosodic features such as fundamental frequency, intensity, sound break (Juncture), spectrogram, pitch wave and intonational contour and so on, respectively, during sarcastic target utterances and baseline utterances.

---

* Intermediate phrase: is a prosodic unit smaller than intonational phrase and greater than the word. The border of the intermediate phrase in the tone tier is displayed by the phrasal tone.

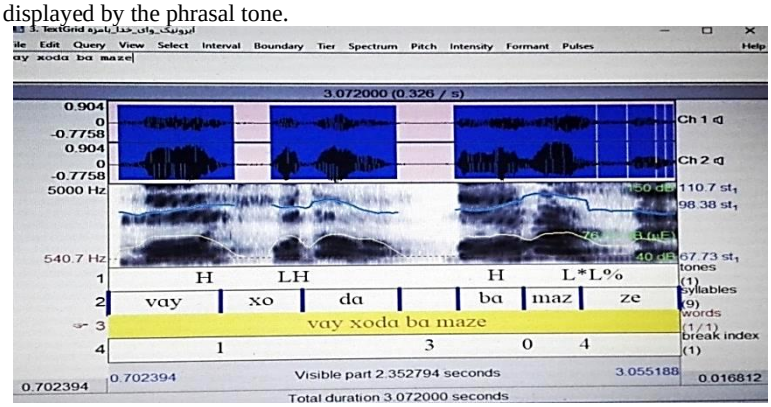


Figure 4. Wave shape, intonational contours and different tiers of TOBI system in the sarcastic utterance: "Oh my God! Cutie!"

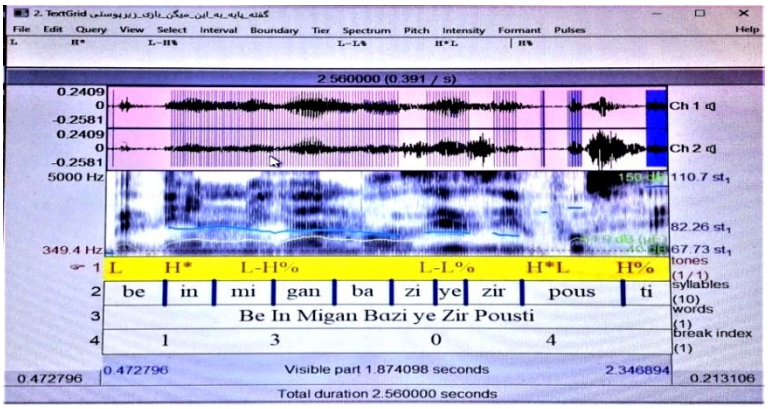


Figure 5. Wave shape, intonational contours and different tiers of TOBI system in baseline utterance: "This is called a subtle acting!"

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

5.4. statistical method

After collecting the data, the researcher entered them in the Excel program and performed statistical analysis using SPSS statistical program. In order to compare between the utterances and also control and investigate the effects of irony subtypes on the quantitative values of prosodic elements in ironic utterances was used T-Tests and a MANOVA statistical test with irony type as an independent variable and the 4 acoustic dimensions specified in this study as dependent variables, respectively. Then, Chi-Square tests were used to examine the independence of these variables from the "utterance type" variable (composed of the measures of "sarcastic target utterances" and "non-sarcastic baseline utterances"). In this research, all the performed tests were non-invasive and sampling of the subjects was carried out with obtaining their informed consent.

The results of T-Tests, which were performed to determine the independence of the means with "utterance type" as an independent variable and 4 acoustic dimensions as dependent variables indicated that only MSD (Mean Syllable Duration) values between baseline utterances and target sarcastic utterances were significantly different at $p<0.05$ (see Table 2 in the section below).

Table2. Mean and Standard Deviation values of 4 acoustic dimensions across the 48 baseline and sarcastic target utterances; F0 and F0 variability values in semitones (St.); intensity values are given in decibels (dB); and MSD values are given in milliseconds (ms).

Acoustic dimension	Baseline utterances		Sarcastic utterances	
	M	SD	M	**SD**
F0 (semitone)	87.57	4.93	89.50	**5.54**
F0 variability	24.68	16.53	39.70	**26/81**
Intensity (dB)	71.60	3.70	73.08	**4.10**
Syllable duration (ms)	216	8.33	315*	**14.55**

Note. F0 = fundamental frequency (pitch); F0 variability = F0 standard deviation (pitch variation respect to F0 mean); intensity = amplitude MSD = mean syllable duration. All semitone values are relative to 1 Hz. Significance '*' = $p < 0.05$

In order to check the potential effects of irony subtypes on prosodic measurements of ironic utterances was used a repeated-measures MANOVA, with "irony subtype" as an independent variable and 4 acoustic measures as dependent variables. As expected, no effect of "irony subtypes" was observed on any of the 4 audio dimensions of ironic utterances. The overall pattern was not significant, $F = 1/24,\ p = ./41\ (\eta^2 = ./23)$.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

4.6. Voice Quality

The results of the voice quality analysis indicated that, whereas 51% of the ironic utterances were produced with a non-modal voice quality, only 14% of the baseline utterances were produced with a falsetto (sharp and unnatural) or creaky voice. The results of a Chi-square test demonstrated that the presence of voice quality features was significantly different between baseline and ironic utterances, $(x^2(1) = 8.32\ ,\ P < 0.05$.

4.7. Juncture or Pause

A juncture is a slight pause that appears between words or syllables when speaking to indicate the end of a word or syllable and the beginning of another word or syllable. Sometimes the juncture or the pause alone may not be sufficient to express the meaning and we need other phonetic elements to convey the meaning. In most cases, in addition to the pause, the intonation is also involved in determining the meaning. In the present study, the acoustic pause parameter was also investigated in these two speech styles. The results from this study showed a significant difference in the mean values of acoustic degree between these two speech styles. In ironic speech, the participants spoke with virtually more pauses than in their own usual speech. The mean of acoustic pause degree was 34.71% for ironic utterances and 14.58% for baseline utterances, respectively.

Discussion

Multiple t-test analyses revealed that the mean absolute values of the cues applied across the 48 ironic and baseline utterances were significant. In the table below, the mean measures of the absolute value from two lexical-syntactic and prosodic markers are shown separately for the baseline and ironic target utterances.

Table3. Percentage values of the mean absolute value of the lexical-syntactic, prosodic and visual cues extracted from the baseline and ironic target utterances.

Cue type	**Mean absolute value**	
	Ironic target utterance	**baseline utterance**
lexical-syntactic	1.77	**0.68**
prosodic	4.36	**2.18**

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

It is worth noting that in the present study, regardless of the pragmatic strategy (i.e. irony subtype) applied by the speaker, ironic utterances had a mean of 8.93 in prosodic cues (i.e. 4.36) compared to a mean of 4.78 in baseline utterances (i.e. 2.18). In practical terms, It means that utterances consistently marked with prosodic cues, with at least 5 auditory strategies. On the contrary, the mean absolute value of lexical-syntactic cues was 0/68% for non-ironic utterances and 1.77% for ironic target utterances. If we compare the mean absolute value of lexical-syntactic cues to the mean absolute value of prosodic cues, the concentration of prosodic signals is higher than lexical-syntactic markers. That is, whereas the mean absolute value of prosodic cues was 2.18 for baseline vs. 4.36 for ironic utterances, the mean absolute value of lexical-syntactic cues was 0.68 for baseline and 1.77 for ironic utterances.

Broadly speaking, the results from the present study demonstrate that 1) Persian speakers produce 5 types of irony based upon Gibbs proposed classification (2000) through a varied set of lexico-syntactic and prosodic cues; 2) The results from comparing fundamental frequency and its variability range in both ironic and baseline speech styles indicated that in spontaneous discourse conditions, Persian speakers talked to each other in lower voice to express their ironic intent. The frequency variability range in ironic speech was also higher than baseline. In addition, the high pitch range, on the one hand, gave the speech of the speakers an exaggerating state and gave rise to attract the audience attention and on the other hand, due to covering more notes, the speaker's speech is perceived to be more melodious. As a result, the audience will express more social responses, and in this way, the main purpose of sarcastic speech, i.e. conveying a negative intention with a positive expression, has been achieved. From an educational viewpoint, these attentive and communicative effects will help the development of language and learning by directing the audience's attention to linguistic inputs; 3) The ultimate probing indicated that the prosodic cues, in production task of Persian irony, operate together as a single system and contribute to intelligibility and detection of an ironic speech even in the absence of verbal markers. Furthermore, there were no correlated patterns between certain types of irony and certain pitch patterns / intonational contours. 4) From relevance theory perspective, on the one hand, we divide an interpersonal communication into two categories, verbal and non-verbal communication, which in this study both of them were assessed through a multi-facets process, and paralinguistic is one of those vocal and non-verbal aspects of speech related to how to say things in special situations, that

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

is, the tone of speech but not related to what is said. Therefore this issue conducts the listener to detect and perceive the speaker's meaning, and on the other hand, whereas the stimulus materials employed in this experiment were a selection of satirical drama series, the results obtained from the investigation of acoustic correlates irony in the spontaneous discourse conditions can be applied in the scope of drama and performance in the genres of irony, sarcasm or satire in the direction of a more natural and professional performance in the audio setting department. 5) During the study and exact observation of the video recordings made via the experiment of irony production, the researcher found the presence and function of visual cues in directing the listener to perceive the ironic meaning of the speaker by conveying his/her attitude or feelings (through facial expressions, smiling, laughing or head movements) and also in revealing the speakers desire to test the perception of the ironic utterance (by conducting his/her gaze towards the listener). Therefore, it seems that conducting a production and perception experiment with a specific purpose to test the perceptual communication of gesture signals with perceiving irony is efficient in realizing this purpose and can be the focus of future.

Acknowledgments

The present study is extracted from the thesis of the doctoral course in Linguistics. The authors of the study consider it necessary to express their sincere gratitude to the research officials and counseling center of Azad University of Firuzabad for all the support and providing the laboratory and tools for this experimental study, and the professors of linguists and psychologists and students who assisted us to conduct, assess and improve the quality of this research.

Ethics statement

All ethical principles are considered in this article. The participants were informed of the purpose of the research and its implementation stages. Principles of the Helsinki Convention were also observed.

Author contributions

All authors contributed to the study conception and design, material preparation, data collection, and analysis. All authors contributed to the article and approved the submitted version.

Funding

The authors did (not) receive support from any organization for the submitted work.

Conflict of interest

This research did not receive any grant from funding agencies in the government, public, commercial, or non-profit sectors could be construed as a potential conflict of interest

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

References

Adachi, T. (1996). “Sarcasm in Japanese,” Studies in Language19, 1–36.

Attardo, S., Eisterhold, J., Hay, J. & Poggi, I. (2003). “Multimodal markers of irony and sarcasm”. Humor: International Journal of Humor Research, 16(2): 243-260.

Bryant, G. & Fox Tree, J. (2002). “Recognizing Verbal Irony in Spontaneous Speech”. Metaphor and Symbol, 17(2): 99-117.

Bryant, G. A. (2005). “Is there an Ironic Tone of Voice?” Language and speech, 48: 257-277.

Bryant, G. A. (2010). *Prosodic contrasts in ironic speech*. Discourse Processes 47 (7), 545-566.

Capelli, C. A., Nakagawa, N. and Madden, C. M. (1990). “How children understand sarcasm: the role of context and intonation”. Child Development, 61, 1824–1841.

Cheang, H.S. & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. https://doi.org/10.1121/1.3177275

Cheang, H.S. & Pell, M. D. (2008). “The sound of sarcasm”. Speech Communication, 50(5): 366-381.

Culpeper, J. (1996). Towards an anatomy of impoliteness. Journal of pragmatics, 25, 349-367.

Cutler, A. (1974). On saying what you mean without meaning what you say. In: LaGaly, M.W., Fox, R.A., Bruck, A. (eds.). Chicago Linguistic Society, Chicago, IL, pp. 117-127.

Cutler, A. (1976) Spoken language: dynamic and changing. Journal de Psychologie Normale et Pathologique 73 (3-4), pp. 273-304.

Cutler, A. (1976). Beyond parsing and lexical look-up: an enriched description of auditory sentence comprehension. In: Wales, R.J., Walker, E. (eds.). Dik, S.C., Kooij, J.G. (eds.), North Holland Linguistic Series, vol. 30, pp. 133-149. North Holland Publishing Company, Amsterdam.

Fónagy, I. (1971). “Synthèse de l’ironie. Analyse par la synthèse de l'intonation motivée”. Phonetica 23(1) : 42–51.

Gibbs, R. W. (2000). *Irony in talk among friends*. Metaphor and Symbol 15, 5-27.

Haiman, J. (1998). “Talk is cheap: Sarcasm, alienation, and the evolution of language”. USA: Oxford University Press.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]

Haverkate, H. (1990). A speech act analysis of irony. Journal of PragmaticsVolume 14, Issue 1, February 1990, Pages 77-109. https://doi.org/10.1016/0378-2166(90)90065-L

Ladd, D. R. (1978). The structure of intonational meaning: Evidence from English. Indiana University press: Bloomington and London, 198, 1980. Reviewed by Jill House, University College London.

Ladd, D. R. (1996/2008). "Intonational Phonology". Cambridge: Cambridge University Press.

Laval, V. and Bert-Erboul, A. (2005). "French-speaking children's understanding of sarcasm: The role of intonation and context". Journal of Speech Language and Hearing Research, 48(3): 610-620.

Leech, G. N. (1983). Principles of Pragmatics. London: Longman.

Loevenbruck, H., Ben Jannet, M. A. (2013). Prosodic cues of sarcastic speech in French: slower, higher, wider. 14th Annual Conference of the International Speech Communication Association (Interspeech 2013), Aug 2013, Lyon, France. Proceedings of Interspeech, pp.3537-3541.

Rockwell, P. (2000a). "Lower, slower, louder: Vocal cues of sarcasm," J. Psycholinguist. Res.29, 483–495.

Rockwell, P. (2000b). "Actors', partners', and observers' perceptions of sarcasm," Percept. Mot. Skills 9, 665–668.

Rockwell, P. (2005). "Sarcasm on television talk shows: Determining speaker intent through verbal and nonverbal cues," in Psychology of Moods, edited by A.Clark(Nova Science, New York), pp. 109–140.

Sharrer, L. and Christman, U. (2011). "Voice Modulations in German Ironic Speech". Language & Speech, 54(4): 435-465.

Voyer, D. & Techentin, C. (2010). Subjective Auditory Features of Sarcasm. Journal of Metaphor and Symbol. Philadelphia. Volume 25, 2010 - Issue 4. Pages 227-242. https://doi.org/10.1080/10926488.2010.510927/.

Wichmann, A. (2000) Intonation in Text and Discourse. Pearson Education, London.

Wichmann, A. (2000).The attitudinal effects of prosody, and how they relate to emotion. Journal of speech and emotion. Newcastle, Northern Irland, UK.

Winner, E, and Leekman, S.(1991). "Distinguishing irony from deception: Understanding the speaker's second-order intention," Br. J.Dev.Psychol.9, 257–270.

[Downloaded from ijer.hormozgan.ac.ir on 2026-11-20]

[DOI: 10.22034/3.4.299]