A fine-grained analysis of the acoustic cues involved in verbal irony recognition in French

Santiago González-Fuente [1]*, Pilar Prieto* [2,1]*, Ira Noveck* [3]

[1] Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain
[2] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
[3]Centre National de la Recherche Scientifique, Laboratoire Langage, Cerveau et Cognition (L2C2), Universite de Lyon, Bron, France

santiago.gonzalez@upf.edu, pilar.prieto@upf.edu, noveck@isc.cnrs.fr

Abstract

Research on verbal irony has found that prosodic features such as pitch range expansion, syllable lengthening, and specific intonational contours are common prosodic resources that languages use to mark irony in speech. Yet little is known about the relative weight of these prosodic features in the detection of irony in languages that use these three prosodic correlates. In this paper we present the results of two experiments designed to shed light on the relative contribution of the acoustic cues involved in verbal irony detection. The first experiment –a production task- was designed to confirm that these three prosodic features are typical of irony in French. Indeed, we found that these three features revealed themselves when readers produced a just read ironic utterance as opposed to a literal one. The second experiment presented the same stories as Experiment 1's, but removed a single word from the context that would otherwise determine whether an acoustically presented utterance (which had been based on a literal reading) was ironic or not. The last word in these utterances was manipulated synthetically so as to create five experimental conditions: Not Modified, Modified Pitch Range, Modified Duration, Modified Intonation, and Modified All. Results showed (a) that speakers tended to interpret utterances as ironic when all acoustic modifications (i.e. pitch range expansion, syllable lengthening and marked intonation) were presented together (i.e. Modified_All); and (b) that the Modified_Duration and Modified_Intonation conditions were significantly more likely to encourage ironic readings than the Not_Modified and Modified_Pitch_Range conditions.

Index Terms: verbal irony recognition, acoustic correlates of verbal irony, prosody, intonation, French.

1. Introduction

Verbal irony is a form of non-literal language in which a speaker produces an utterance whose actual meaning is at variance with or even opposite to its literal verbal content [1, 2]. For an ironic meaning to be understood, listeners have to detect the discrepancy between the speaker's words and his/her intentions [3, 4]. Research on the perception of irony has shown that its detection is based on a complex cognitive process that needs to take into account contextual, lexical, gestural, and prosodic factors [e.g., 5, 6]. Despite the fact that irony does not necessarily have to be cued with overt linguistic marking and can be successfully interpreted by relying only on contextual cues [7], several studies have shown that verbal irony is commonly produced with lexical, gestural, and prosodic marks working together [e.g., 8, 9] which helps listeners detect a speaker's ironic attitude with respect to the proposition expressed [9, 10, 11]. Regarding prosodic cues, production and perception studies have shown that speakers of different languages employ overall prosodic modulations to mark ironic speech (e.g., pitch raising or lowering, higher pitch variability, intensity raising, and utterance lengthening), and that listeners rely on these prosodic features to detect an ironic intention [e.g., 2, 5, 12, 13, 14].

However, results on the perceptual value of these prosodic cues are not consistent across studies and across languages. Whereas utterance lengthening has been documented as one of the prosodic cues that signal the presence of ironic intent across languages, the results of pitch and intensity are more controversial, as they show inconsistencies in their relative importance and also in their specific patterns (i.e., higher vs. lower), both across studies and across languages. Moreover, little is known about the potential role of specific 'ironic' pitch contours and their relationship with other cues. As Loevenbruk et al. [15] note, more data is needed in order to determine whether specific intonational contours (with pitch accents at specific locations and/or specific boundary tones) can affect the detection of irony and even override the presence of other prosodic features. To our knowledge, there are no previous studies focusing on the perception of verbal irony that specifically investigate the extent to which pitch range expansion, syllable lengthening, and specific intonational contours influence an ironic interpretation.

Here we report on the results of two experiments on the expression and perception of irony in French. A recent study on sarcasm perception [15] showed that sarcastic productions in this language are characterized by utterance lengthening and by increased F0 modulations, and also by a slight final rise instead of the falling contour. Our first experiment involved a production task designed to confirm the typical overall prosodic features of ironic utterances in French, as well as the potential use of specific intonational patterns. In our second experiment, a perception task was carried out to test for the relative contribution to verbal irony perception of the different prosodic marks found in the production task. Importantly, we performed all these prosodic modulations separately on critical

nuclear words (e.g. the word "*fantastic*" in the ironic utterance "Your new haircut is *fantastic*") [13, 16].

2. Experiment 1. Production task

2.1. Methods

2.1.1. Materials

In this experiment we used twelve 'story-frameworks' adapted from Spotorno et al. 2012 [17]. These stories led to a target sentence whose reading could be ironic or literal depending on one key word or phrase in the context. Thus, a negative context followed by a positive sentence was intended to induce an ironic interpretation of the target sentence, whereas a positive context followed by the same positive sentence was intended to trigger a literal interpretation of the target sentence. For example, depending on whether a story about two friends watching a film at the cinema is presented in a negative context (e.g., the film is boring) or a positive one (e.g., the film is thrilling), the utterance in (1) thus trigger will receive either an ironic or a literal interpretation.

(1) "*Nous sommes allées voir un film formidable*"
("We went to see a great film")

All 12 story-frameworks were six lines long. The first three sentences described a common everyday situation and an exchange between two characters who know each other only casually (we wished to avoid situations involving closely-related interlocutors); the fourth and fifth sentences described the outcome of the situation (positive in the literal version and negative in the ironic version); and the sixth line contained the target sentence, in all cases a comment uttered by one of the two characters. All target sentences were between 10 and 12 syllables long. Importantly, the target sentence was exactly the same in both ironic and literal versions and the key word to determine whether the sentence was ironic or not was always the last word of the utterance (e.g., "formidable" in (1)). In order to control for pragmatic variables that could affect the interpretation of the target sentence, in all the story-frameworks the main characters were females. Crucially, we also controlled for irony subtype by ensuring that all the ironic situations consisted of a negative context followed by a positive sentence that could be clearly recognized as a critical (i.e., sarcastic) remark about the situation.

In order to prevent negative contexts from being cues to the presence of ironic remarks, we included in the experimental design two stories in which a negative context (e.g. two opera singers performing terribly) led to a banal non-ironic remark (e.g., "We will do better the next time."). Following Spotorno et al. [17], we call these two preliminary stories *decoys*. We also included two stories called *fillers* in which a positive context (e.g., successful rhinoplasty surgery) was followed by a banal remark (e.g. "I'm very happy with my new nose"). Participants were thus randomly presented with a total of 16 stories: 6 ironic, 6 literal, 2 decoys, and 2 fillers. The experiment was carried out in French.

2.1.2. Participants, procedure, and data analysis

Ten women (mean age 21.3; SD 3.2) participated in Experiment 1. The 10 participants were each seated at a table in front of a laptop. The experimental materials were presented to them in the form of a Microsoft PowerPoint 2010 slide show. They were instructed to read all 16 six-line stories carefully, line by line, by clicking the space bar on the keyboard, and to read the last sentence of the story (always a comment in quotation marks) out loud as if they were the character in the story. Before the experiment *per se* started, they were presented with three example stories to ensure they had understood the procedure. All sessions were recorded in a soundproof room using a PMD660 Marantz professional portable digital recorder. The recorded target utterances were analyzed using PRAAT software [18] in terms of their prosodic characteristics (average F0, F0 variability, duration, average intensity, and intonation pattern—i.e., tonal nuclear configuration) at two different levels, namely across the whole target sentence and during the last critical word of the sentence.

2.2. Results

A total of 60 ironic performances and 60 literal performances were submitted to acoustic analysis at the two levels noted above, i.e., across the whole sentence and during the last word of the target sentence.

2.2.1. Target sentence

All the scores of average F0, F0 variability (both measured in semitones), average intensity (in dB) and duration (in ms.) were higher in ironic utterances compared to literal utterances. Four Generalized Linear Mixed Model (GLMM) tests were run using IBM SPSS Statistics 20.0 [19]. The variable UTTERANCETYPE (ironic vs. literal) was set as the fixed factor, and the dependent variables were the four acoustic dimensions. Subject and utterance were set as random factors. The results showed that only average F0 ($p < 0.05$) and duration ($p < 0.01$) were significantly higher in ironic compared to literal renditions. The mean difference in average F0 between literal and ironic sentences was 0.92 semitones. The duration of ironic utterances was on average 26.39% longer than the duration of literal utterances. Finally, though F0 variability showed a near significant effect between conditions ($p = 0.059$), this was not the case for average intensity ($p = 0.11$).

2.2.2. Last word of target sentence

The last word of the sentence was also subjected to acoustic analysis, and the results were very similar to those found at the overall utterance level. Mean scores for all acoustic correlates were higher in ironic utterances compared to literal utterances, and GLMM tests revealed a significant difference between words produced in literal vs. ironic conditions in terms of Average F0 ($p < 0.01$), Duration ($p < 0.01$) and in this case also F0 Variability ($p < 0.05$). The mean difference in Average F0 between literal and ironic sentences was 1.01 semitones and ironic utterances exceeded literal utterances in duration on average by 29.57%. It was also observed that 23% of final words were produced with a specific pitch accent (H+!H*) and 28% with a final mid boundary tone !H% (see [15]). All in all, the results of Experiment 1 confirm the same pattern of results reported in previous literature for the prosodic parameters of ironic utterances at the sentence level. Moreover, we identified a specific pitch configuration associated with ironic speech in French, namely H+!H* !H%.

3. Experiment 2. Perception task

On the basis of the results of Experiment 1, we carried out a second experiment in which the recordings of the target critical words were manipulated so that they differed in terms of pitch range, duration, and intonation pattern.

3.1. Methods

3.1.1. Materials

Experiment 2 used the same 12 stories ending with the same target sentences as Experiment 1, but in this case, the fourth line of the story was presented with a missing word (crucially, the key word that rendered the story either literal or ironic, see Example 2), while the concluding comment in line six was absent from the written text but presented auditorily.

(2) Cynthia and Léa sing together in the same opera group.
On the night of the premiere they meet at the theater.
The show begins exactly on time.
Their performance is ########.
After the show, Cynthia says to Lea:
"Our performance was **formidable**." (Presented in audio)

To obtain the audio target sentences for Experiment 2, 16 women (mean age 23.2; SD 4.1) were asked to read the same 12 sentences used in Experiment 1 so that a different voice was heard for each story. In order to avoid the presence of other emotional acoustic cues, we presented the test sentences without any context. The sentences were recorded in a soundproof room using a PMD660 Marantz professional portable digital recorder. For each of the 12 target sentences, we prepared 5 different versions by modifying the acoustic parameters of the last critical word (e.g., *formidable* in (2)). The five acoustic modifications were performed separately in accordance with the results of Experiment 1, as follows: (1) with no modification; (2) with pitch modification (expanding the range of the pitch accent by 1 semitone); (3) with duration modification (a 30% increase for each word); (4) with intonation pattern modification (from L*L% to H+!H* !H%); and (5) with all modifications (pitch, duration, and intonation pattern) (see Figure 1).
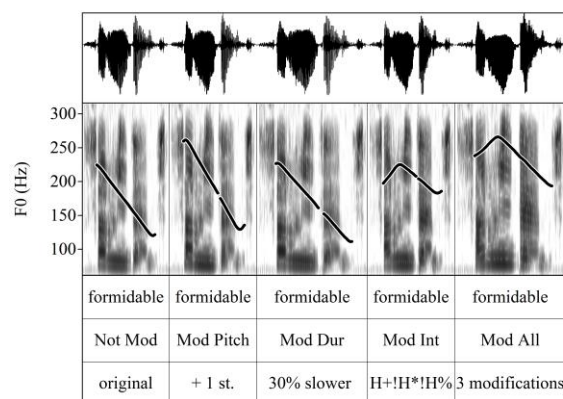

Figure 1. Example of the 5 ACOUSTICMODIFICATION conditions performed on the word 'formidable' in the sentence *Nous sommes allées voir un film formidable* 'We have seen a fantastic film'.

As in Experiment 1, we incorporated two stories in which a negative context led to a banal non-ironic remark (i.e., *decoys*) and two other stories in which a positive context was followed by a positive remark (i.e., *fillers*) (see section 2.1.1).

3.1.2. Participants and procedure

The experimental design was a between-subjects design with three different groups and ACOUSTICMODIFICATION as the independent variable. Thus, for one group of subjects (33 participants; 14 male, 19 female) the 16-story questionnaire (6 ironic, 6 literal, 2 decoys, and 2 fillers) contained conditions (1), (2), and (5); for a second group (37 participants; 15 male, 22 female), the questionnaire contained conditions (1), (3), and (5); and for the last group (31 participants;11 male; 19 female) the questionnaire contained conditions (1), (4), and (5). Moreover, to obtain responses for all possible story-condition combinations, each group of subjects was divided into three subgroups which were presented with different versions of the questionnaire (9 versions in total, 3 for each group of subjects). A total of 101 French native speakers (mean age 32.3; SD 12.6) participated individually in this perception task, which was prepared with Survey Gizmo [20] and carried out online.

After completing a brief questionnaire about their age, educational level, and native language, participants were randomly presented with the 16 stories. They were instructed to read each story carefully, listen to the audio file (which contained the last sentence of the story), and then propose what they thought would be the most appropriate word to fill the blank at the end of the fourth line (signaled with a string of '#'s; see (2)). Importantly, they were forewarned that some, all, or none of the stories might have an ironic reading. This forewarning was deemed necessary because in prior piloting trials it was found that if the possibility of an ironic reading was not mentioned, test items never received an ironic reading except for a few in the 'Modified_All' condition. While this observation is interesting in itself for irony studies, the main objective of this experiment was to determine whether individual acoustic modifications provide cues to irony.

3.2. Results

A total of 1212 responses were obtained and classified as ironic/non-ironic/ambiguous by a native French speaker with the responses classified as ambiguous then being eliminated from subsequent analysis. Importantly, the classifier of responses was blind to the prosodic condition. The remaining 877 responses (with values '0' indicating non-ironic and '1' ironic) were subjected to a Generalized Linear Mixed Model (GLMM) test using IBM SPSS Statistics 20.0 [19]. The variable RESPONSE was set as the dependent variable. The fixed factor was ACOUSTICMODIFICATION (5 levels: 'Not_Modified', 'Mod_Pitch_Range', 'Mod_Duration', 'Mod_Intonation', 'Mod_All'), and subject, age, educational level, experimental group, experimental group, story-framework, and gender were set as random factors. Results revealed a main effect of ACOUSTICMODIFICATION ($F(4,872) = 65.673$, $p<.000$). Figure 2 shows the mean rate of ironic scores (from 0 'literal' to 1 'ironic') depending on the ACOUSTICMODIFICATION condition. The graphs show that the 'Modified_All' condition triggers the highest scores (near 0.8).

In all the other conditions the perception of irony is under 0.45, and in 'Not_Modified' the score is 0.21.
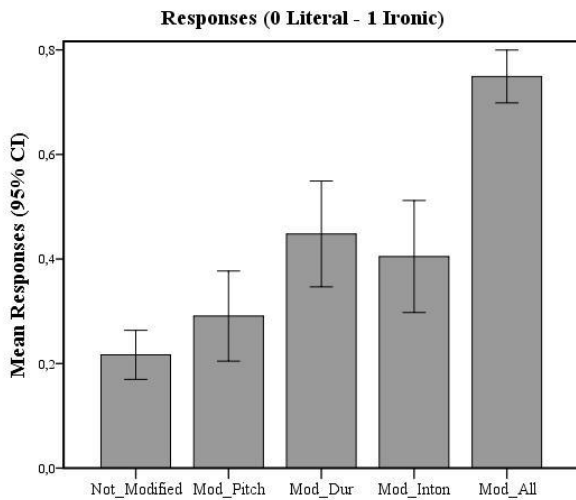

Figure 2. Mean rate of ironic scores (from 0 'Literal' to 1 'Ironic', y-axis) for ACOUSTICMODIFICATION conditions ('Not_Modified', 'Mod_Pitch_Range', 'Mod_Duration', 'Mod_Intonation', 'Mod_All', x-axis).

This significant main effect was further submitted to Bonferroni post-hoc analyses by pairwise comparisons, which interestingly revealed no significant differences between 'Not_Modified' and 'Modified_Pitch_Range' conditions (and between 'Modified_Duration' and 'Modified_Intonation' (see Figure 3).

In sum, the results show that (a) participants detected irony when the target word was modified in three different acoustic parameters; (b) expanding the pitch range of the pitch accent seems not to be a powerful acoustic cue of irony; and (c) word lengthening and a specific intonation pattern (H+!H% !H%) seem to be more related to the expression of irony than pitch range expansion.
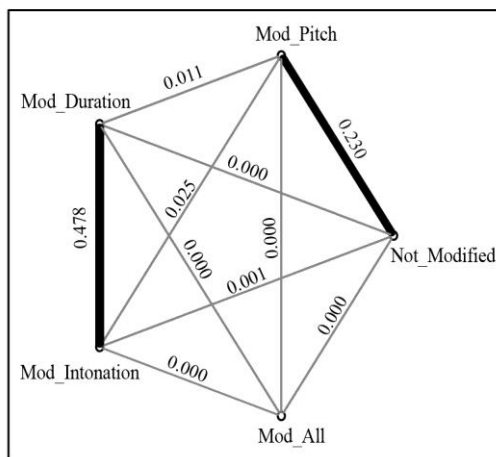

Figure 3. Results of pairwise comparison Bonferroni post-hoc analyses. Thin lines show significant and bold lines non-significant statistical differences between ACOUSTICMODIFICATION conditions. P-values for each pairwise comparison are displayed on the lines.

4. Discussion and Conclusions

The results of Experiment 1 are consistent with previous research on verbal irony detection in French [15, 21], in terms of both overall sentence analysis and analysis of the critical word. A significant difference between ironic and literal utterances was found in F0 mean (higher for ironic utterances), F0 variability (higher), and target duration (longer durations). Moreover, the results also showed that 23% of the ironic utterances were produced using a specific intonational contour (H+H!*H%), which has been described in the French_ToBI annotation system as containing a specific pragmatic meaning related to disagreement and the expression of counterfactual statements, that is, utterances made when the speaker thinks that the listener holds a contrasting view [22].

The results of Experiment 2 showed that speakers judge as ironic those utterances containing all the prosodic cues that have been reported to be important in previous studies on verbal irony perception [e.g., 1, 3, 5, 8, 9, 18]. Specifically, the study showed that French speakers are more able to detect irony when all the characteristic acoustic features (pitch range expansion, syllable lengthening, and marked intonation) are presented together. Moreover, duration and intonation cues were found to be more powerful for irony detection than pitch range expansion. The fact that pitch modifications are not as relevant for irony detection as word lengthening might be due to the multifunctionality of pitch variation in language. Previous research on emotional prosody has found that speakers of different languages detect emotions mainly by pitch variations, and not by duration modulations [23, 24, 25]. We think that the pitch characteristics present in ironic utterances may be more closely related to the expression of affective meaning than to any specific pragmatic meaning, unlike intonational contours. Pitch accent expansions may signal emphasis, but, as our fine-grained analysis shows, listeners judged sentences with this emphatic pitch accent to be no more effective than those with no additional prosodic markings.

Overall, our results confirm previous findings reported in the literature on the production of irony in French and also shed some light on the potential crosslinguistic weight of different acoustic markings in verbal irony detection. However, further studies with other languages on the detection of the ironic tone of voice are needed in order to confirm the existence of these pragmatic tendencies.

5. Acknowledgments

6. References

[1] G.A. Bryant, "Is verbal irony special?," *Language and Linguistics Compass*, vol. 6, no. 11, pp. 673–685, 2012.

[2] H.S. Cheang and M.D. Pell "The sound of sarcasm," *Speech Communication*, vol. 50, pp. 366–381, 2008.

[3] R.W. Gibbs, *The poetics of mind: figurative thought, language, and understanding*. New York: Cambridge University Press, 1994.

[4] P.M. Pexman, "It's fascinating research: The cognition of verbal irony," *Current Directions in Psychological Science*, vol. 17, pp. 286–290, 2008.

[5] P. Rockwell, "Lower, slower, louder: Vocal cues of sarcasm," *Journal of Psycholinguistic Research*, vol. 29, pp. 483–495, 2000.

[6] G.A. Bryant, "Verbal irony in the wild," *Pragmatics and Cognition*, vol. 19, no. 2, pp. 291-309, 2011.

[7] S.L. Ivanko and P.M. Pexman, "Context incongruity and irony processing," *Discourse Processes*, vol. 35, pp. 241–279, 2003.

[8] D. Wilson and T. Wharton, "Relevance and prosody," *Journal of Pragmatics*, vol. 38, pp. 1557-1579, 2006.

[9] S. González-Fuente, V. Escandell-Vidal, and P. Prieto, "Gestural codas pave the way to the understanding of verbal irony," *Journal of Pragmatics*, vol. 90, pp. 26–47, 2015.

[10] V. Escandell-Vidal and M. Leonetti, "Fronting and Irony in Spanish," in A. Dufter and A. Octavio de Toledo (Eds.), *Left Sentence Peripheries in Spanish: Diachronic, Variationist and Typological Perspectives*, vol. 8, John Benjamins, Amsterdam, pp. 309–342, 2014.

[11] J. House, "Intonation structures and pragmatic interpretation," in S. Ramsaran (Ed.), *Studies in the pronunciation of English*, Routledge, London, pp. 38–57, 1990.

[12] G.A. Bryant, "Prosodic contrasts in ironic speech," *Discourse Processes*, vol. 47, no. 7, pp. 545–566, 2010.

[13] X. Padilla, "¿Existen rasgos prosódicos objetivos en los enunciados irónicos?," *Oralia*, vol. 14, pp. 203–224, 2012.

[14] L. Scharrer, U. Christmann, and M. Knoll, "Voice Modulations in German Ironic Speech," *Language and Speech*, vol. 54, no. 4, 435–465, 2011.

[15] H. Loevenbruck, M. BenJannet, M. D'Imperio, M. Spini, and M. Champagne-Lavau, "Prosodic cues of sarcasm speech in French: slower, higher, wider," in *INTERSPEECH 2013 – 14th Annual Conference of the International Speech Communication Association, August 2013, Lyon, France, Proceedings*, 2013, pp. 3537–3541.

[16] S. Attardo, J. Eisterhold J, J. Hay, and I. Poggi, "Multimodal markers of irony and sarcasm," *International Journal of Humor Research*, vol. 16, pp. 243–260, 2003.

[17] N. Spotorno, E. Koun, J. Prado, J.B. Van Der Henst, and I.A. Noveck, "Neural evidence that utterance-processing entails mentalizing: The case of irony," *NeuroImage*, vol. 63, pp. 25–39, 2012.

[18] P. Boersma and D. Weenink, Praat: doing phonetics by computer. Version 5.3.23. Computer program. Available at http://www.praat.org/. Retrieved July 2012.

[19] IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

[20] C. Vanek, S. McDaniel. SurveyGizmo Online Survey Software. Available from: http://www.surveygizmo.com. Retrieved December 2014.

[21] V. Laval and A. Bert-Erboul, "French-speaking children's understanding of sarcasm: The role of intonation and context," *Journal of Speech Language and Hearing Research*, vol. 48, no. 3, pp. 610–620, 2005.

[22] E. Delais-Roussarie, B. Post, M. Avanzi, C. Buthke, A. Di Cristo, I. Feldhausen, S. Jun, P. Martin, T. Meisenburg, A. Rialland, R. Sichel-Bazin, and H. Yoo, "Developing a ToBI System for French," in S. Frota and P. Prieto (Eds.), *Intonational variation in Romance*, Oxford University Press, Oxford, pp. 63–100, 2015.

[23] B. Alinezhad and E. Vaysi, "An Acoustic Study of Emotivity Prosody Interface in Persian Speech Using the Tilt Model," *Journal of Teaching Language Skills*, vol. 3, no. 2, pp. 1–27. 2011.

[24] R. Altrov, "Aspects of Cultural Communication in Recognizing Emotions," *Trames. Journal of the Humanities and Social Sciences*, vol. 17, no. 2, pp. 159–170, 2013.

[25] L. Anolli, L. Wang, F: Mantovani, A. De Toni, "The voice of emotion in Chinese and Italian young adults," *Journal of Cross-Cultural Psychology*, vol. 39, no. 5, pp. 565–598, 2008.