Asymmetry of affect in verbal irony understanding: What about the N400 and P600 components?

Stéphanie Caillies, Pamela Gobin, Alexandre Obert, Sarah Terrien, Alexandre Coutté, Galina Iakimova, Chrystel Besche-Richard

Asymmetry of affect in verbal irony understanding: What about the N400 and P600 components?

Stéphanie Caillies[1], Pamela Gobin[1,2], Alexandre Obert[1,2], Sarah Terrien[1], Alexandre Coutté[3,4], Galina Iakimova[3], Chrystel Besche-Richard[1]

[1]Laboratoire C2S (Cognition, Santé, Société) EA6291, Université de Reims Champagne-Ardenne, Reims, France

[2] Pôle Universitaire de Psychiatrie, CHU de Reims, EPSMM, Reims, France

[3]LAPCOS (Laboratoire d'Anthropologie et de Psychologie Clinique, Cognitive et Sociale) EA7278, Université Côte d'Azur, Nice, France

[4]CeRSM (Centre de Recherches sur le Sport et le Mouvement) EA2931, Université Paris Nanterre, France

Asymmetry of affect in verbal irony understanding: What about the N400-P600 components?

Abstract

We investigated the neurocognitive processes behind the asymmetry of affect observed in irony understanding, where ironic criticism is more easily understood than ironic praise. We recorded the ERPs of participants while they listened to positive (e.g., "These children are always smiling") or negative (e.g., "His son is very unfortunate") remarks pronounced with a sincere or ironic prosody. Participants had to decide whether or not the speaker was sincere. Behavioural results confirmed the asymmetry of affect phenomenon and ERP results revealed that the N400 and P600 were differentially sensitive to the negative or positive emotional connotations of the speaker's messages. These findings shed new light on the cognitive processes behind biphasic N400/P600 cycles, and how they are differentially affected by negativity.

Keywords: ironic criticism; ironic praise; emotional connotation; ERPs

1. Introduction

Verbal irony is a form of nonliteral language in which speakers can convey implied meanings that intentionally contradict the literal meanings (e.g., saying "He is bright!" about somebody who says stupid thing). When they use verbal irony, speakers simultaneously communicate literal information and a dissociative attitude toward this literal information, and thus toward the referent (event, person or object). This attitude can be manifested by counterfactual aspects of the context, or by the speaker's facial cues, body posture, or tone of voice (Gibbs & Colston, 2007). For example, "That's a really good idea!" can be taken as either a compliment or a criticism, depending on the tone of voice, referred to as *prosody* in the literature.

Although there is no consensus as to which vocal cues are most indicative of verbal irony, research has confirmed that prosody can inform listeners about a speaker's ironic attitude. For example, Bryant and Fox Tree (2002) found that participants could correctly distinguish between spontaneously produced ironic versus nonironic utterances based on prosody. Consistent with this, Rockwell (2000) showed that sarcastic irony can be detected when no context is provided, further emphasising the role of prosody in irony perception (see also Voyer & Techentin, 2010). Prosodic aspects of speech can therefore supplement or modify the meaning of the spoken sentence by providing valuable clues to the speaker's attitude. At least two kinds of prosody have been identified in the literature: emotional prosody and attitudinal prosody (Mitchell & Ross, 2013). While *emotional prosody* refers to the emotions conveyed by the tone of voice, *attitudinal prosody* refers to the expression of a person's attitude towards an event, person or object, be it scepticism, doubt, enthusiasm, or boredom (Mitchell & Ross, 2013; Wickens & Perry, 2015). Attitudinal prosody can drastically modify the literal meaning of verbal irony, by indicating the speaker's dissociative attitude, although emotional prosody can also be used (Bryant, 2010). For example, a speaker can pronounce "That's just

great" with an ironic intent and using an angry prosodic contour. In the present study, we therefore sought to investigate further the understanding of ironic intention based on the attitudinal prosody of utterances presented in isolation.

According to several theories, verbal irony is generally used to express a negative attitude towards a person or event (Kumon-Nakamura, Glucksberg, & Brown, 1995; Sperber & Wilson, 1981). An ironic sentence therefore takes the form of a positive sentence used to convey a negative meaning, as in the sarcastic comment "He is bright!" about somebody who says something wrong. However, less common, but still heard, is ironic praise, in which a negatively worded utterance conveys a positive message, as when someone says "You have a hard life!" to a friend going on a long vacation (Schwoebel, Dews, Winner, & Srinivas, 2000). The literature shows that there is an asymmetry in the processing of these two kinds of irony, in that ironic criticism is understood sooner and more easily than ironic praise (Hancock, Purdy, & Dunham, 2000; Harris & Pexman, 2003). Clark and Gerrig (1984) called this effect the *asymmetry of affect*.

The present study specifically addressed this asymmetry of affect by asking whether the nature of the neurocognitive processes underlying the understanding of a speaker's ironic intention depends on the positive or negative connotation of the ironic remark. We examined these processes via event-related brain potentials (ERPs).

Two ERP components are of particular interest when studying language comprehension: the N400 and the P600. Brouwer, Fitz, and Hoeks (2012) even assumed that language comprehension proceeds in biphasic N400/P600 cycles. The N400 is a negative component that occurs between 300 and 500 ms, with a peak at about 400 ms after stimulus onset, and is characterised by a centroparietal distribution. It was initially interpreted as being sensitive to semantic violations (Kutas & Hillyard, 1980), but later found to be sensitive to semantic features at both word (Holcomb & Neville, 1990; Kutas & Hillyard, 1989) and

sentence levels (Kutas & Federmeier, 2007). N400 amplitude was originally assumed to reflect semantic integration (Osterhout & Holcomb, 1992; Van Berkum, Hagoort & Brown, 1999), but more recently a semantic retrieval hypothesis has gained ground (Kutas & Federmeier, 2000; Van Berkum, 2009). According to this hypothesis, the N400 is a general index of the difficulty of retrieving stored conceptual knowledge associated with a meaningful stimulus, which is dependent on both the stored representation itself and the retrieval cues provided by the preceding context (Van Berkum, 2009). Brouwer et al. (2012) went as far as to postulate that the N400 component is modulated by the retrieval of lexical-semantic information, rather than by semantic integration or any other kind of compositional semantic processing. ERP findings on irony processing are not consistent regarding the N400, even though they all relate to ironic criticism. While some studies have reported a greater N400 amplitude for ironic statements than for literal ones (Katz, Blasko, & Kazmerski, 2004), others have not (Balconi & Amenta, 2008; Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013). One possible explanation for this disparity of results in the literature is that the N400 can be modulated by the task demand and by the degree of familiarity of the ironic comment. For example, Cornejol et al. (2007) found an increased N400 for irony when participants were required to apply a holistic interpretative strategy ("if you can understand what the character means by the remark"), but not when they applied an analytic strategy focusing on the congruency of the final sentence of the story, thus supporting the semantic integration hypothesis. For their part, Filik, Leuthold, Wallington, and Page (2014) observed a less negative waveform in the N400 time window for unfamiliar ironic remarks than for literal ones, but there was no difference when the ironic remarks were familiar, consistent with the hypothesis that N400 modulation reflects semantic retrieval.

The P600 (also known as the late positive component, LPC[1]) is commonly observed in the 500-900 ms time window, with a parietal topography. The cognitive processes reflected by this component were initially interpreted as involving the manipulation of syntactic information (Hagoort, Brown, & Groothusen, 1993), but were subsequently associated with the notion of conflict monitoring (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg, 2007; Van Petten & Luka, 2012). Brouwer et al. (2012) postulated that P600 amplitude reflects integration compositional processes vested in interpretation. They further suggested that, rather than reflecting a single process, it subsumes many different subprocesses that can be differentiated and labelled on the basis of differences in onset, duration and scalp distribution. Consistent with an interpretative integration view, the P600 effect has frequently been observed in ERP studies of irony comprehension. For example, when Spotorno et al. (2013) exposed participants to sentences preceded by a substantial context biasing to either an ironic or a literal interpretation, they found that ironic sentences induced greater P600 amplitude than literal ones. Similarly, Regel, Gunter, and Friederici (2011) observed a P600 effect for irony processing regardless of modality (visually or auditory) and task (comprehension or passive reading). Regel, Coulson, and Gunter (2010) replicated the irony-related P600-like effect for visually presented materials, and demonstrated its dependency on the participant's knowledge about a character's typical (ironic vs. nonironic) communicative behaviour. Along the same lines, Filik et al. (2014) observed that both familiar and unfamiliar auditory ironic sentences elicited a greater P600 than literal sentences.

Although these ERP studies investigated the neurocognitive processes underlying the understanding of ironic versus literal statements, none of them focused on the asymmetry of affect. As a reminder, *asymmetry of affect* refers to the fact that ironic criticism ("He is bright!"), which is negatively connoted, is understood sooner and more easily than ironic praise

("You have a hard life"), which is positively connoted. The main aim of the present study was to examine how the neurocognitive processes underlying the understanding of irony, exclusively expressed through dissociative prosody, are modulated by the emotional connotation of the ironic remark (i.e., ironic criticism vs. praise).

We recorded the ERPs of 49 participants while they listened to positive (e.g., "These children are always smiling") or negative (e.g., "His son is very unfortunate") remarks pronounced with a sincere or ironic prosody. Participants had to decide whether or not the speaker was sincere. To understand the ironic remark presented in isolation and the speaker's underlying ironic intention, listeners have to integrate two divergent sources: the meaning of the expression and its intonation. Based on the findings of irony studies with a similar design and the requirement that participants adopt a holistic approach (Cornejol et al., 2007), we expected the N400 to vary in amplitude as a function of prosody, with greater amplitude for an ironic versus sincere intonation, reflecting either the greater integrative cost of irony or the greater difficulty of retrieving stored knowledge associated with the words making up the ironic expression. Furthermore, given the asymmetry of affect, modulation might vary according to the emotional connotation of the ironic remark, with greater N400 amplitude for ironic praise versus criticism, reflecting the greater processing cost of noncanonical versus canonical irony. Regarding the P600, based on what has been repeatedly observed in the irony literature, we would expect to find a greater amplitude for ironic versus sincere prosody. The overall pattern of N400 and P600 modulations in the different conditions would shed light on the processes involved in the asymmetry of affect.

2. Method

2.1. Participants

Participants were 49 students from Reims University (37 women; mean age = 24.0 years, *SD* = 5.33, range = 18-42). They were all native French speakers, had normal or corrected-to-normal

vision, and reported no history of psychiatric or neurological disorders. The study was approved by the relevant French ethics committee, and all participants gave their written informed consent after receiving a full description of the study.

2.2. Material

The experimental material consisted of 80 sentences (40 negative and 40 positive) created especially for the present study. All the sentences took the same form (subject/verb/adjective), where the adjective determined the emotional valence. The verb used was "to be", with the exception of two cases for which the verb was "to seem". Examples of positive or negative sentences, as per the valence of the adjective, are provided in Table 1.

The emotional valence of the adjectives had been established in a previous study in which 204 native French speakers, none of whom took part in the ERP study, rated 103 adjectives on a 7-point emotional scale ranging from 1 (*negative*) to 7 (*positive*). Based on these ratings, we selected 40 highly positive (mean = 5.92, *SD* = 0.48) and 40 highly negative (mean = 2.21, *SD* = 0.43) adjectives, which were inserted as final words in our experimental sentences. It should be noted that the positive and negative adjectives had similar mean lexical frequencies of 11.92 (*SD* = 22.45) and 11.99 (*SD* = 22.9), $F(1, 39) < 1$ (lexique.org database; New, Pallier, & Ferrand, 2005).

In order to assess other potential differences between the material conditions, we asked independent participants to perform two tasks, one assessing the cloze probability of the adjectives, the other assessing the familiarity of the sentences. In the cloze-probability task, 191 participants aged 18-48 years (*M* = 20.81, *SD* = 5.28; 22.5% men) were shown a list of sentences minus the last word (i.e., adjective) in an online test, and were asked to type a word that plausibly completed the sentence. Sentences were divided into two lists of 40 sentences each. Participants were randomly assigned to one of the two lists. Results revealed that cloze probability was close to 0 for most sentences, and there was no significant difference between

sentences ending with a positive adjective (.003) and sentences ending with a negative one (.008) ($W = 762$, $p = .586$). In the familiarity task, inspired by Filik et al. (2014)'s methodology, 78 students aged 19-35 ($M = 21.25$, $SD = 2.31$; 16.7% men), were asked to indicate how familiar they were with each sentence used ironically on a 7-point Likert-like scale ranging from 1 (*Not familiar*) to 7 (*Familiar*). They were randomly assigned to one of two lists of 80 sentences each (40 experimental sentences and 40 fillers). Results indicated that sentences ending with a positive adjective (3.69) were rated as more ironically familiar than those ending with a negative one (2.33), $t(78) = 8.53$, $p < .001$.

Each of the 90 experimental sentences was pronounced several times by a professional male actor who was a French native speaker, with both a sincere and an ironic intention. All utterances were digitally recorded and saved as individual audio files. Four judges evaluated the different voice recordings. Only consensual voice recordings were selected. The experimental material thus consisted of 40 negative sentences (ending with a negative adjective) and 40 positive sentences (ending with a positive adjective), each pronounced with a sincere and an ironic intonation (see Table 1). The 160 pronounced sentences were divided into two lists, so that participants saw 40 positive and 40 negative sentences but each only in one version, sincere or ironic. It should be noted that in the case of the ironic intonation, the initial positive sentence "Their children are always smiling" was meant to be understood as negative (criticism), and the negative sentence "The spectators were disappointed" as positive (praise).

Table 1 about here

2.3. Acoustic analysis

The purpose of the acoustic analysis was to determine whether, for the same speech segment, the ironic prosody differed clearly from the sincere prosody. The analysis was carried out using PRAAT (Version 6.0. 2.0.; www.praat.org). Prosodies were assessed in terms of segment duration, fundamental frequency, and intensity. These three variables were chosen because they

are strongly associated with perceptually recognisable prosodic qualities (Voyer & Techentin, 2010). The results of the acoustic analysis are set out in Table 2.

Table 2 about here

We ran an analysis of variance (ANOVA) on each of the three measures. Consistent with previous results (Laval & Bert-Erboul, 2005; Regel, Gunter, & Friederici, 2011), ironic prosody was characterised by a longer duration, $F(1, 156) = 233$, $p < .001$, higher pitch (mean fundamental frequency), $F(1, 154) = 270.10$, $p < .001$, and lower intensity, $F(1, 156) = 78.8$, $p < .001$) than sincere prosody. Whatever the prosody, negative and positive segments significantly differed solely in duration, $F(1, 156) = 5.73$, $p < .02$, with positive sentences (ending with a positive adjective) being longer than negative ones (ending with a negative adjective). These results confirmed that ironic prosody differed in its perceptual features from neutral prosody.

2.4. Procedure

Participants were tested individually, seated in front of a monitor at a distance of nearly 60 cm. They listened to either the list 1 or the list 2 of the materiel, each containing 94 sentences. Each trial began with the display of a fixation cross for 1000 ms. Then the auditory sentence was played during the presentation of a black screen for a variable duration. This was followed by a fixation cross for 500 ms and by a question mark inviting the participant to answer the question "Does the speaker think what he says?" The time between the display of the question mark and the following trial was fixed at 4000 ms. Before the experimental task, participants underwent a training session featuring six sentences that were not used in the experimental task.

2.5. Electroencephalographic recording and ERP analysis

Electroencephalograms (EEGs) were recorded using a 10-20 system electrode cap with 32 channels, connected to an amplifier (BrainAmp; Brain Products, Munich, Germany). The reference electrode was located midway between Fz and Cz, and the ground electrode between

Fz and the prefrontal electrodes. Impedance was kept below 5 kΩ for all participants. The amplification gain was 1000. We recorded the electro- from two electrodes derived from FP1 and FP2 placed on and below the outer canthus of the right eye. EEG data were filtered with bandpass cut-offs of 0.1-100 Hz and a 30-Hz low-pass filter. A notch 50-Hz filter was also applied. Segmentation was then applied from 200 ms before the target onset to 1500 ms after. We performed ICA decomposition, and electro-oculographic and other linear artefacts were identified using SASICA (Chaumont, Bishop, & Bush, 2015). The ERP signal was then re-referenced offline to the digital mean of the left and right mastoids. We carried out baseline correction for the 200-ms epoch before target onset. Segments containing residual artefacts were then checked visually and using the moving window peak-to-peak threshold function. A moving 200-ms window with a 100-ms step and 100-μV processing threshold were set. This led to the rejection of 4.79% of the data. The mean number and range of averaged trials were equivalent across all four conditions: negative adjective + ironic prosody (13.6 ± 4.20), positive adjective + ironic prosody (16.43 ± 3.26), negative adjective + sincere prosody (18.3 ± 2.10), and positive adjective + sincere prosody (16.1 ± 3.50).

Based on visual inspection and on previous ERP studies, we selected the following nine electrodes for the N400 (300-500-ms time window) and P600 (500-900-ms time window) statistical analyses: F3, Fz, F4, C3, Cz, C4, P3, Pz and P4. We also conducted exploratory analyses for the 900-1500 ms time window (positive post-P600 component). These analyses were justified by both the visual inspection of the waveforms and previous observations of late positive effects of emotionally impacting stimuli in similar time windows (e.g., Benning et al., 2016; Thiruchselvama, Blecherta, Sheppesa, Rydstromb, & Gross, 2011).

2.6. Statistical analysis

For the behavioural data, the error rate was submitted to generalised linear mixed models, because of the nature of the dependent variable, via restricted maximum likelihood estimation

(see Table 3 for comparative and absolute fit statistics). First, the null model (Model 0), including participants – crossed with adjective valence and prosody –, and items – crossed with prosody –, as random terms to consider the specific design of our experiment (cf. Barr, Levy, Scheepers, & Tily, 2013), served as the point of comparison for fit statistics presented in Table 3. In the following step (Model 0b), we tested the effect of familiarity, as it differed significantly between our conditions. Then, the effect of prosody (ironic *vs*. sincere – within-participant factor) was tested (Model 1), before adding the main effect of adjective valence (positive *vs*. negative – within participant factor) (Model 2). Finally, we added to the model our particular effect of interest: the interaction between prosody and adjective valence (Model 3).

Concerning the ERP data, mean amplitudes for the N400, P600 and post-P600 components were separately submitted to linear mixed models (see Table 3 for detailed fit statistics). For each component, the null model (Model N0/Model P0/Model L0) integrated the two random effects structures (i.e., variance across participants, crossed with adjective valence and prosody, and items, crossed with prosody), and Model 1 (Model N1/Model P1/Model L1), the effect of familiarity. Then, the respective effects of prosody and adjective valence as well as their interaction were directly modelled in the following step (Model N2/Model P2/Model L2), as they were the three effects of interest. Statistics for each effect in these models were presented in Results section.

All statistical analyses were performed using R software (R Core Team, 2017), as well as the lme4 (Bates, Maechler, Bolker, & Walker, 2014), MuMIn and phia packages.

3. Results

Table 3 about here

3.1. Behavioural data

The effect of prosody on the error rate was reliable, indicating that participants made more errors when the prosody was ironic (0.24) than when it was sincere (0.11). The effect of

adjective valence was not reliable. There also was a significant Prosody x Adjective valence interaction. As shown in Figure 1, this interaction showed that the difference in the error rate between the ironic and sincere prosodies was significant for sentences ending with a negative adjective (ironic praise) (d = 0.21), χ^2 = 57.46, p < .001, but not for sentences ending with a positive one (ironic criticism) (d = -0.01), χ^2 = 0.0003, p = .987. These behavioural results are consistent with the asymmetry of affect (Clark & Gerrig, 1984) observed in the literature: negative sentences with an ironic prosody (ironic praise) were harder to understand than positive sentences with an ironic prosody (ironic criticism).

Figure 1 about here

3.2. ERP data

ERP waveforms for the nine electrodes are shown in Figure 2, along with topographical images.

Figure 2 about here

300-500 ms - N400 component

The results of Model N2 revealed a significant effect of the Adjective valence x Prosody interaction, F(1, 27257.9) = 17.43, p < .001. For sentences ending with a negative adjective, mean N400 amplitude was greater for ironic prosody (i.e., ironic praise; -2.070 µV) than for sincere prosody (-1.579 µV), whereas for ones ending with a positive adjective, mean N400 amplitude was greater for sincere prosody (-1.408 µV) than for ironic prosody (i.e., ironic criticism; -0.894 µV). No other significant effect was observed (all ps > .10).

500-900 ms - P600 component

The results of Model P2 showed that the Adjective valence x Prosody interaction was also reliable, F(1, 27247.2) = 23.88, p < .001. For sentences ending with a negative adjective, mean P600 mean amplitude was less positive for ironic prosody (i.e., ironic praise; 0.172 µV) than for sincere prosody (1.034 µV), whereas for ones ending with a positive adjective, mean P600

amplitude was larger for ironic prosody (i.e., ironic criticism; 1.253 µV) than for sincere prosody (0.823 µV). No other significant effect was observed (all *p*s > .10).

900-1500 ms – post-P600 component

The results of Model L2 showed a significant effect of prosody, $F(1, 46.8) = 5.02$, $p = .030$, with a post-P600 component amplitude that was greater for ironic prosody (2.832 µV) than for sincere prosody (1.887 µV). Moreover, the Adjective valence x Prosody interaction was significant, $F(1, 27234.7) = 6.06$, $p = .014$. The effect of prosody on post-P600 component amplitude was greater for sentences ending with a positive adjective (ironic prosody: 3.101 µV vs. sincere prosody: 1.764 µV; $d = 1.337$ µV) than for ones ending with a negative adjective (ironic prosody: 2.563 µV vs. sincere prosody: 2.144 µV; $d = 0.419$ µV). No other significant effect was observed (all *p*s > .10).

4. Discussion

The main goal of the current study was to investigate the neurocognitive processes behind the asymmetry of affect generally observed in irony comprehension. First, our behavioural data confirmed the existence of this asymmetry, as a sentence ending with a positive adjective pronounced with an ironic prosody (ironic criticism) allowed for a better understanding of what the speaker meant than a sentence ending with a negative adjective that was ironically pronounced (ironic praise). Thus, it was easier to infer what the speaker meant when he ironically said “This audience is very pleasant” than when he ironically said “The spectators were disappointed”. These two types of irony (i.e., criticism and praise) both involve mock evaluations of circumstances with a valence that contrasts with the speaker’s true appraisal, but are apparently sustained by different processes or knowledge. What do the N400 and P600 tell us about the processes responsible for this asymmetry?

ERP data did not show a modulation of the N400 according to either prosody or valence, but consistent with the asymmetry of affect, this component was modulated by their interaction. For sentences ending with a negative adjective, mean N400 amplitude was greater for ironic prosody (i.e., ironic praise) than for sincere prosody, whereas for ones ending with a positive adjective, mean N400 amplitude was greater for sincere prosody than for ironic prosody (i.e., ironic criticism). In other words, prosody differently affected the N400 as a function of emotional information. There are two possible interpretations, based on the main theories regarding the N400. From an integrative viewpoint (Kutas & Federmeier, 2000), we can assume that sentence-prosody inconsistency impeded meaning construction more when the adjective's valence was negative rather than positive. From a retrieval viewpoint (Kutas, Van Petten, & Kluender, 2006), we can argue that the inconsistency impeded the retrieval of negative adjectives from memory more than that of positive adjectives from memory. It should be noted that our results did not appear to be compatible with the prediction derived from Brouwer et al. (2012)'s model, which posits that the N400 component is modulated by the retrieval of lexical-semantic information, rather than by semantic integration or any other kind of compositional semantic processing. Moreover, the observed effect of irony on the N400 appears consistent with the finding of a recent study (Baptista, Manfredi, & Boggio, 2017) in which ERP responses coupled with transcranial direct current stimulation (tDCS) were recorded during the visual presentation of verbal irony remarks, criticism and praise following a picture describing a counterfactual context. This study also demonstrated an asymmetry of affect, regardless of tDCS stimulation, with greater negativity in response to ironic language than in response to literal language, but only in the praise condition. Taken together, these results indicate that the N400 can be functionally modulated by the emotional connotation of ironic remarks, whether they are presented aurally or visually. Thus, our findings show that the

N400 can be modulated by emotional information in addition to the task demand (Cornejol et al., 2007) or degree of familiarity (Filk et al., 2014).

Results regarding the P600 confirmed previous ERP findings showing that irony processing induces an enhancement of the P600 component (Filik, Leuthold, Wallington, & Page, 2014; Regel, Gunter, & Friederici, 2011; Spotorno, Cheylus, Van Der Henst, & Noveck, 2013), but only for canonical irony (i.e., ironic criticism). Consistent with the asymmetry of affect, results indicated that the P600 was modulated by the emotional connotation resulting from adjective valence and prosody, but contrary to what we observed for the N400 component, it was ironic criticism rather than ironic praise that elicited a greater P600. These results confirm that the P600 effect is the most recurrent outcome of pragmatic phenomena, reflecting the inference of the speaker's meaning. In addition to canonical irony, P600 effects are observed during the comprehension of metaphors (Bambini, Bertini, Schaeken, Stella, & Di Russo, 2016; Obert, Gierski, & Caillies, 2018), idioms (frontal topography; Canal, Pesciarelli, Vespignani, Molinaro, & Cacciari, 2017), and presuppositions (Domaneschi, Canal, Masia, Lombardi, & Bambini, 2018). However, the fact that the P600 was sensitive to ironic criticism, but not to ironic praise, merits further discussion. It supports one of the hypotheses put forward by Regel et al. (2011), whereby the P600 effect is a reflection of the processing of emotional information conveyed by canonical irony, insofar as it mainly expresses the speaker's disappointment at an event (saying things as they should be). We return to this point below.

The overall pattern of N400 and P600 modulations according to type of irony (criticism vs. praise) was quite complex. We did not observe greater N400 modulation for ironic criticism than for its literal counterpart, but greater P600 modulation, and found the opposite pattern for ironic praise. This differential effect of emotional information on N400 and P600 component modulations raises questions about the processes involved. In the irony literature, two main

explanations have been put forward for the asymmetry of affect. Ironic praise may be harder to interpret because although ironic criticism tends to implicitly allude to conventions about politeness and saying nice things, ironic praise does not (Kumon-Nakamura, Glucksberg, & Brown, 1995). Within this framework, ironic praise (i.e., a sentence with a negative valence but a figurative positive meaning), which is literally negative, is a less effective reminder of our implicit positive expectations or norms, and is consequently harder to process than ironic criticism. Another possible explanation for this asymmetry of affect is that ironic praise involves a double negation that is more difficult to process. According to Giora (1995), to understand ironic praise, the negative literal meaning has to be negated. However, as the ironic praise sentences in our study did not contain any negation markers (just negative adjectives), this theory appears less relevant here. Turning to allusional pretence theory (Kumon-Nakamura et al., 1995), it can be argued that ironic praise is more incongruous than irony criticism because it is not usual to voice (literal) evaluative negative thoughts, making it hard to process such counternormative remarks and find figurative interpretations.

Looking afresh at our ERP results with this theory in mind, the greater N400 modulation we observed for ironic praise (i.e., a sentence with a negatively valenced adjective but a positive figurative meaning) could result from the unconventional feature of its inherent literal negative remark, namely saying bad things to people. Likewise, the greater P600 modulation observe for ironic criticism (i.e., a sentence with a positively valenced adjective but a negative figurative meaning) could reflect the fact that the figurative negative connotation was harder to process than the figurative positive connotation of the ironic praise. It seems that different processes were involved, as reflected by both the N400 and the P600 components, which exhibited asymmetry of affect modulations. It is as if processes, behind each ERP component, were modulated differentially by emotionality in the ironic condition, with a particular focus on the negative meaning. By nature, the negativity is at the level of the

sentence valence for ironic praise while it is at the level of the connotation for ironic criticism. Therefore, N400 was more modulated by the negativity of the adjective while P600 was more affected by the negativity of the connotation. It may therefore be harder to start processing the sentence meaning with the prosody for ironic praise than for ironic criticism, because of the nonconventional emotional feature of the literal remark, as reflected by the greater N400. By contrast, as illustrated by the greater P600, the processing may be more costly for ironic criticism because of its figurative negative connotation.

But how did negativity modulate these ERP components? As we did not include neutral utterances in our material, we could not specifically assess the absolute effect of negativity. Further studies are therefore needed to answer this particular question. Nevertheless, our exploratory analyses for the 900-1500-ms window, based on visual exploration of the ERPs, gave us some additional information, specifically on the post-P600 component. Considering the LPC in its broadest sense, components occurring in late time windows have been shown to respond, among others, to the emotional content of verbal stimuli, and more specifically to their emotional valence. Their amplitudes usually appear larger for valenced stimuli than for neutral ones, although results are not consistent regarding a positive or negative advantage (for a review, see Citron, 2012). When we scrutinize our results, we can see that the post-P600 varied according to the asymmetry of affect. More specifically, we only found a greater modulation of this component for an ironic remark than for its literal counterpart when the former was negative (i.e., ironic criticism). In the literature, authors suggest that a greater LPC reflects the enhanced motivation and arousal that is experienced in response to affective stimuli (Benning et al., 2016). With this in mind, the post-P600 effect we observed could be linked to the greater emotional impact of ironic criticism, compared with ironic praise.

What are the cognitive processes behind the ERP components involved in irony understanding? Our proposal regarding a negativity bias raises the question of the expectation

of negativity at both semantic and pragmatic levels. Current psycholinguistic theories emphasize the importance of prediction for language understanding (for a review, see Ferreira & Chantavarin, 2018). Processing is viewed as fundamentally Bayesian, with people processing current information in the light of their expectations, leading them to predict the following word. In this framework, both the N400 and the P600 may be triggered by disconfirmation of active predictions, each at a specific level. As it is not usual to voice evaluative negative thoughts, we can assume that people predict hearing a literal positive adjective during irony, or a positive connotation when prosody and sentence converge. However, the fact that cloze probability did not differ between positive sentences (ending with a positive adjective) and negative sentences (ending with a negative adjective) in our study means that this interpretation cannot be confirmed. The functional distinction proposed by Hagoort, Baggio, and Willems (2009) could shed an interesting light on our results. These authors distinguished integration from unification processes, and suggested that integration allows for information to converge to a common memory representation, whereas semantic unification is a constructive process in which a semantic (or pragmatic) representation is built – and not retrieved. Based on this model, we can hypothesise that the sentence-prosody inconsistency of the ironic remark impeded integration more when the adjective was negatively rather than positively valenced, as reflected by the N400. Furthermore, it entailed more unification when the figurative meaning was negative, as indexed by the P600. In this framework, the greater N400 we observed for ironic praise may have reflected more difficult convergence to a common memory representation, owing to the negativity of the sentence, while the greater P600 for ironic criticism may have reflected a higher cost of unification, owing to the negativity of the figurative meaning. Finally, the post-P600 effect may an indicator of the final impact of irony understanding, insofar as it is a rhetorical device used to criticize while saying nice things.

5. Conclusion

Our study suggests that the processes underlying the understanding of a speaker's ironic intention depend on the emotional connotation of the ironic remark (praise vs. criticism). It sheds light on the behavioural asymmetry of affect, improving our understanding of the processes involved, as reflected by the biphasic N400/P600 component modulation, in addition to the post-P600 positive component, and highlighting the sensitivity of these ERP components to the emotional information conveyed by speakers' messages. Further studies including neutral sentences are needed to confirm our finding that N400 and P600 amplitudes are differentially affected by emotional information, and more generally to explore the impact of emotional connotation on ERPs (see Fischler & Bradley, 2006).

Acknowledgments

We would like to thank Fabien Gierski and Flavien Thuaire for their help.

FUNDING

This work was supported by the French National Research Agency (ANR) (grant no. ANR- 11-EMCO-001 01).

References

Balconi, M., & Amenta, S. (2008). Isn't it ironic? An analysis on the elaboration of ironic sentences with ERPs. *The Open Applied Linguistics Journal*, *1*(1), 9–17. http://doi.org/10.2174/1874913500801010009

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology*, *7*(May), 1–14. http://doi.org/10.3389/fpsyg.2016.00559

Baptista, N. I., Manfredi, M., & Boggio, P. S. (2018). Medial prefrontal cortex stimulation modulates irony processing as indexed by the N400. *Social Neuroscience*, *13*(4), 495–510. http://doi.org/10.1080/17470919.2017.1356744

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Benning, S.D., Kowac, M., Campbell, A., Miller, S., Hanna, E.K., Damiano, C.R., …, Dichter, G.S. (2016). Late positive potential ERP responses to social and nonsocial stimuli in youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *46*(9), 3068-3077. 10.1007/s10803-016-2845-y

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. http://doi.org/10.1016/j.brainresrev.2008.05.003

Brouwer, H., Fitz, H., & Hoeks, J. C. J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127-143.

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*(7), 545-566. http://doi.org/10.1080/01638530903531972

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*(2), 99–119. http://doi.org/10.1207/S15327868MS1702_2

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of*

Experimental Psychology: Learning Memory and Cognition, *43*(6), 928–943. http://doi.org/10.1037/xlm0000351

Chaumont, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Citron, F. M. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211–226. http://doi.org/10.1016/j.bandl.2011.12.007

Clark, H., & Gerrig, R. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, *113*, 121-126.

Cornejol, C., Simonetti, F., Aldunate, N., Ibáñez, A., López, V., & Melloni, L. (2007). Electrophysiological evidence of different interpretative strategies in irony comprehension. *Journal of Psycholinguistic Research*, *36*(6), 411–430. doi:10.1007/s10936-007-9052-0

Domaneschi, F., Canal, P., Masia, V., Lombardi Vallauri, E., & Bambini, V. (2018). N400 and P600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics*, *45*, 13–35. http://doi.org/10.1016/j.jneuroling.2017.08.002

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: A synthesis of old and new. *Current Directions in Psychological Science*, 1-6. doi.org/10.1177/0963721418794491

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(3), 811-828. doi:10.1037/a0035658

Fischler, L., & Bradley, M. (2006). Event-related potential studies of language and emotion: Words, phrases, and task effects. *Progress in Brain Research*, *156*, 185-203.

Gibbs, R. J., & Colston, H. L. (2007). *Irony in language and thought: A cognitive science reader*. Mahwah, NJ: Lawrence Erlbaum Associates.

Giora, R. (1995). On irony and negation. *Discourse Processes*, *19*, 239–264.

Hagoort, P., Brown, C., & Groothusen, J. T. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.

Hagoort, P., Baggio**,** G., & Willlems, R. M. (2009). Semantic unification. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 819-835). Cambridge, MA: Massachusetts Institute of Technology.

Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development*, *1*(2), 227–248. http://doi.org/10.1207/S15327647JCD010204

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, *36*(3), 147–165. http://doi.org/10.1207/S15326950DP3603_1

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related potentials. *Language and Cognitive Processes*, *5*, 281-312.

Katz, A. N., Blasko, D. G., & Kazmerski, V. (2004). Saying what you don't mean. Social influences on sarcastic language processing. *Current Directions in Psychological Science*, *13*(5), 186–189. doi:10.1111/j.0963-7214.2004.00304.x

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*, 3–21.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49. doi:10.1016/j.brainres.2006.12.063.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. http://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Kutas, M., & Hillyard, S.A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38-49.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp 659-724). New York: Elsevier.

Laval, V., & 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*, 610–621.

Mitchell, R. L. C., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience and Biobehavioral Reviews*, *37*(3), 471–479. http://doi.org/10.1016/j.neubiorev.2013.01.027

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur Internet: Lexique™. *L'Année Psychologique*, *101*, 447-462.

Obert, A., Gierski, F., & Caillies, S. (2018). He catapulted his words from the dais: An ERP investigation of novel verbal metaphors. *Journal of Neurolinguistics*, *47*, 59-70. doi:10.1016/j.jneuroling.2018.02.008

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806. http://doi.org/10.1016/0749-596X(92)90039-Z

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. http://doi.org/10.1016/j.brainres.2009.10.077

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–93. http://doi.org/10.1162/jocn.2010.21411

R Core Team (2017). R: A Language and Environment for Statistical Computing. R foundation for Statistical Computing, Vienna, Austria. <URL: https://www.R-project.org/>.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*(5), 483–495. http://doi.org/10.1023/A:1005120109296

Schwoebel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: Further evidence. *Metaphor and Symbol*, *15*(1), 47–61. http://doi.org/10.1207/S15327868MS151&2_4

Sperber, D., &Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radical pragmatics* (pp. 295–318). New York: Academic Press.

Spotorno, N., Cheylus, A., Henst, J. Van Der, & Noveck, I. A. (2013). What's behind a P600 ? Integration operations during irony processing, *PLOS ONE*, *8*(6), 1–10. http://doi.org/10.1371/journal.pone.0066839

Thiruchselvam, R., Blechert, J., Sheppes, G., Rydstrom, A., & Gross, J. J. (2011). The temporal dynamics of emotion regulation: An EEG study of distraction and reappraisal. *Biological Psychology*, *87*(1), 84-92. doi:10.1016/j.biopsycho.2011.02.009

van Berkum, J. J. A. (2009). The "neuropragmatics" of simple utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.

Van Berkum, J., Hagoort, P., & Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*, 657-671.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. http://doi.org/10.1016/j.ijpsycho.2011.09.015

Voyer, D., & Techentin, C. (2010). Subjective auditory features of sarcasm. *Metaphor and Symbol*, *25*(4), 227–242. http://doi.org/10.1080/10926488.2010.510927

Wickens, S., & Perry, C. (2015). What do you mean by that? An electrophysiological study of emotional and attitudinal prosody. *PLOS ONE*, *10*(7), 1–25. http://doi.org/10.1371/journal.pone.0132947

Footnotes

[1] The precise meaning of the term *late positive components* remains unclear in the language literature. There is some debate as to whether the LPC group encompasses the P3 and P600. As a result, studies assessing the LPC may focus on either broad or narrow time windows. For clarity's sake, we interpreted the LPC as having a broad meaning, encompassing all the positive components that occur after the classic N400 component.

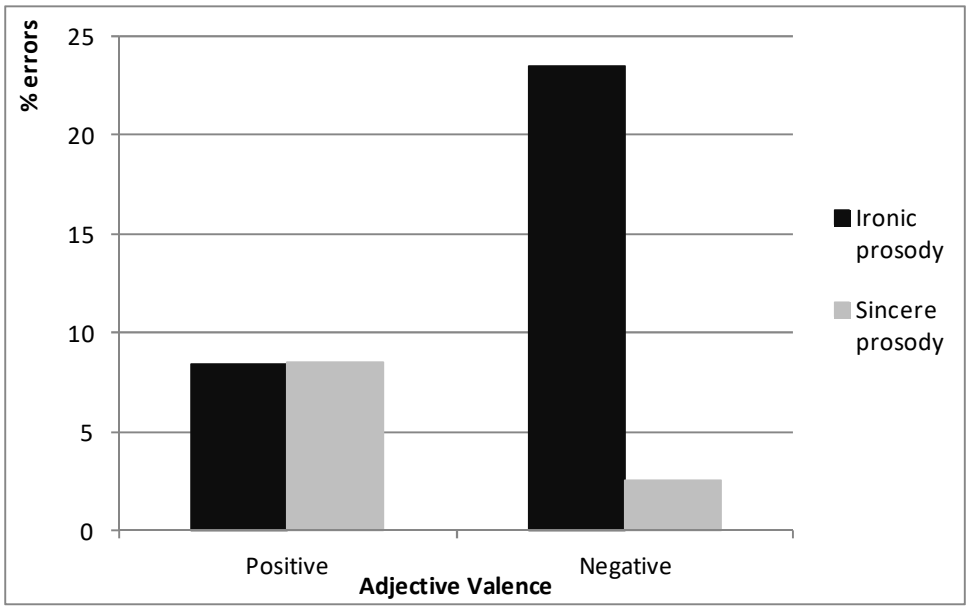


Figure 1. Percentage of errors according to prosody (ironic or sincere) and the valence (positive or negative) of the sentence's final adjective.

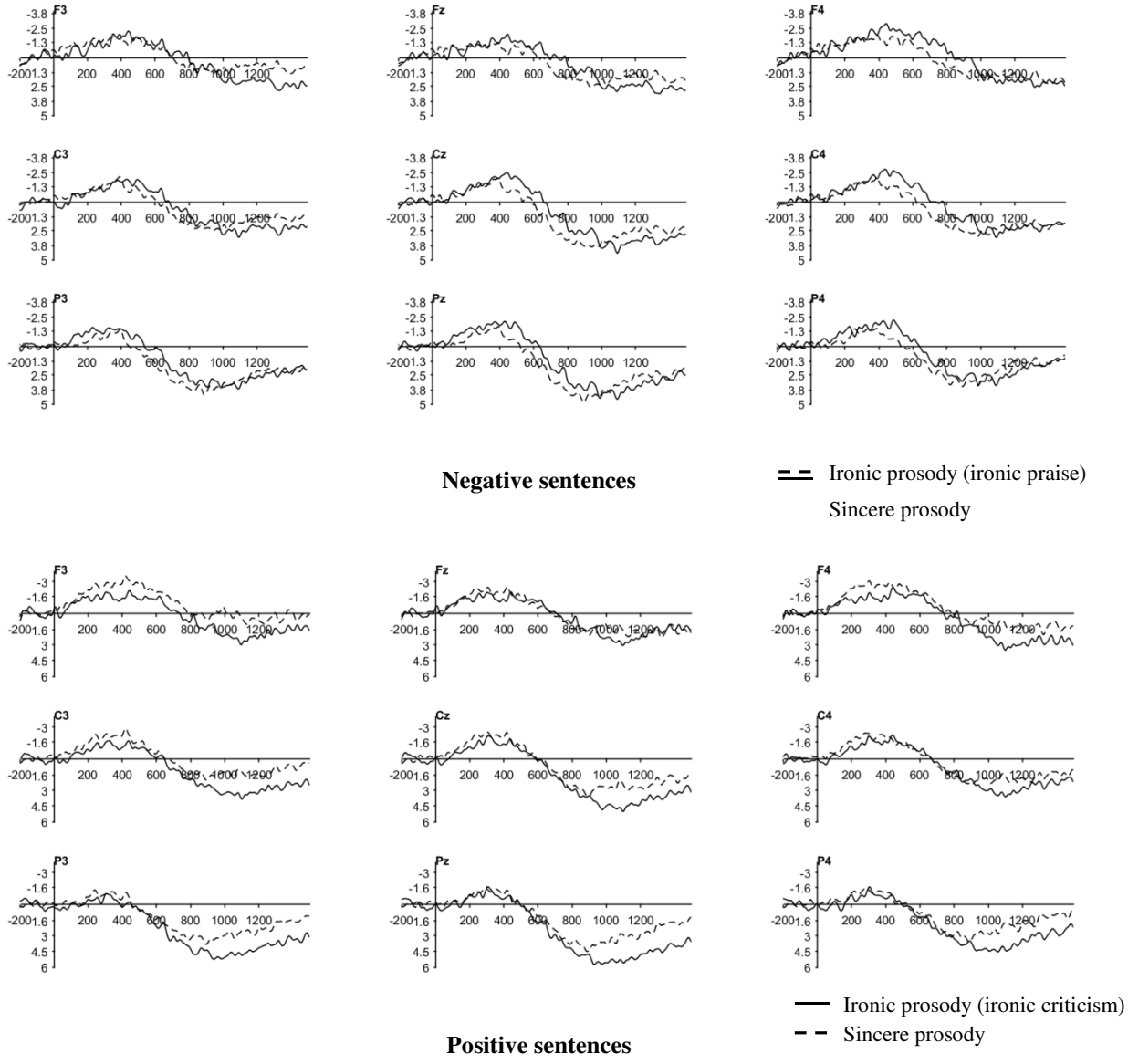
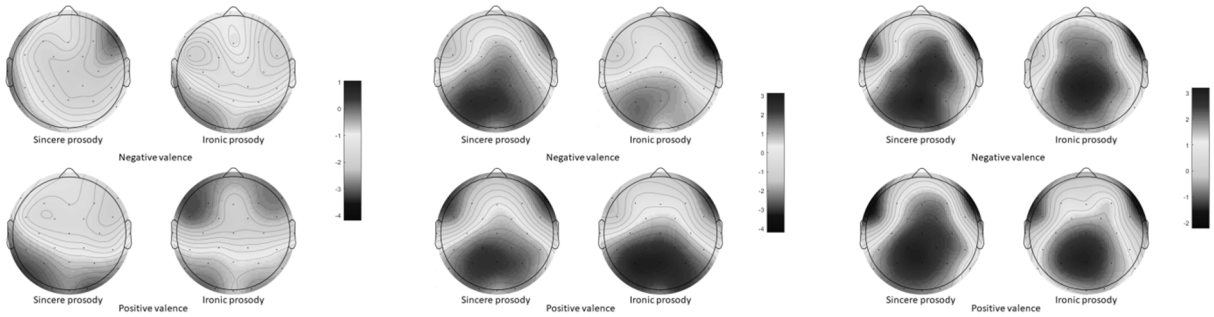


Figure 2a) ERPs for sentences ending with either a negative (top) or a positive (bottom) adjective, as a function of prosody (ironic vs. sincere).



N400 *P600* *Post P600*

Figure 2b) Topographic images for the N400 (left), P600 (centre) and post-P600 (right) amplitudes according to prosody (ironic or sincere) and the valence (negative or positive) of the sentence's final adjective.

Table 1. Examples of sentences ending with a positive or negative adjective

Positive adjective	**Negative adjective**
Ses enfants sont très souriants	Son fils est très malheureux
(Their children are always smiling)	(His son is very sad)
Cet auditoire est très plaisant	Les spectateurs ont été déçus
(This audience is very pleasant)	(The spectators were disappointed)

Table 2. Acoustic analysis of speech segments for differences in duration, fundamental frequency (Hz), and intensity (dB)

Condition	**Valence**	**Mean duration (s)**	**Mean F0**	**Mean intensity**
Ironic	Positive	2.09	149.14	74.16
	Negative	1.44	110.80	77.10
Sincere	Positive	1.90	147.41	74.11
	Negative	1.45	115.85	77.78

Note. F0: fundamental frequency.

Table 3. Fit statistics for the generalised linear mixed models applied to error rates and the linear mixed model applied to N400, P600 and post-P600 amplitude.

	AIC	BIC	LL	R^2	χ^2 (*df*)	*p*
Behavioural data						
Model 0	2591.7	2622.7	-1290.8			
Model 0b	2593.4	2630.6	-1290.7	.372	0.26 (1)	.612
Model 1	2569.5	2612.8	-1277.7	.390	25.98 (1)	$p < .001$
Model 2	2571.2	2620.7	-1277.6	.391	0.313 (1)	.576
Model 3	2541.8	2597.5	-1261.9	.404	31.31 (1)	$p < .001$
N400 amplitude						
Model N0	201773	201830	-100879			
Model N1	201771	201837	-100878	.072	3.48 (1)	$p = .062$
Model N2	201759	201849	-100868	.073	18.79 (3)	$p < .001$
P600 amplitude						
Model P0	206915	206972	-103450			
Model P1	206916	206982	-103450	.101	0.507 (1)	$p = .476$
Model P2	206897	206988	-103438	.102	24.56 (3)	$p < .001$
Post-P600 amplitude						
Model L0	216337	216395	-108162			

Model L1	216337	216403	-108160	.131	2.33 (1)	$p = .127$
Model L2	216331	216421	-108154	.132	12.00 (3)	$p = .007$

Note. Each model is described on page 12.

AIC = Akaike information criterion; BIC = Bayesian information criterion; LL = log likelihood for the model; χ^2 = deviance statistic between the current and previous models.