Analysis of Event-Related Potentials in Sarcasm Comprehension

Xirui He

UC San Diego Cognitive Science Honors Program

Committee: Seana Coulson, Megan Bardolph, Crystal Poole

June 12, 2020

1. Introduction

Sarcasm is classically defined as utterances that are intended to convey the opposite of their literal meaning. This definition could be not precisely correct because the opposition in meaning cannot account for sarcastic such as requests and exclamations. The topic of sarcasm is interesting to cognitive scientists because it provides an opportunity to investigate the relationship between the comprehension of literal and nonliteral meanings in language. For example, when you are standing in a mile-long line in front of a grocery store waiting to pay, you might make a sarcastic statement "*Oh that's great*", while your intended meaning is that it is awful. When hearing the sarcastic statement, the listener would use their knowledge of the context in which the statement is made, and choose to understand if the statement conveys a literal meaning or a sarcastic meaning. We want to investigate in the comprehension process of how sarcastic meaning is decoded by the listener. The question is whether listeners have to decode the literal meaning first and then appreciate the nonliteral one, or whether they can use their knowledge of the context to by-pass the processing of literal meaning and understand the nonliteral meaning directly.

In the example above, when the listener hears the statement "Oh that's great", the listener could first process the literal meaning of the statement (it is a great situation), and by knowing the context that there is a long line to wait for, the listener then access the nonliteral meaning (the situation is awful). On the other hand, by having the understanding of the context, the listener could directly understand the statement in its nonliteral meaning directly (the situation is awful) without thinking about the literal meaning first.

Past research on this topic suggests that there are two main existing models that explain the processing of literal and nonliteral meaning. First, according to Grice and Searle, they predicted that the literal meaning of a sentence is interpreted first. If they encounter information that seems to violate the assumption that speakers cooperate with one another by offering truthful, relevant, and informative information, they then use inferential operations to derive a non-literal intended meaning (Clark, 1979; Clark & Lucy, 1975). Another model, the direct access model, predicts that it is possible to access the contextual relevant meaning of the word

immediately and bypass the literal meaning if there is supportive constraining context (Gibbs, 1994).

Sarcasm is an area of language use that has emotional consequences because speakers use it to both criticize and to compliment one another's behavior. In some contexts, people are more likely to speak sarcastically than others, and the emotional salience of the sarcastic or sincere utterance can be different and have an impact on the interpretation of literal and nonliteral meaning. Consequently, we might be able to learn more about the relationship between literal and nonliteral meanings by looking at when listeners are sensitive to the emotional impact of the literal meaning of individual words compared to the emotional impact of the whole utterance. For example, when you hear someone say "*your child makes me feel so happy/awful*", the words "happy" and "awful" have a strong emotional impact on the listener and they would be the critical words in the sentence that we want to compare.

2. The Present Study

Given our interest in timing of literal and nonliteral meaning processing, the present study used a neural measure with high temporal resolution - event-related brain potentials (ERPs) - to study how the brain activity temporally correlates with sarcasm comprehension. An ERP is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event. It is often used to evaluate brain functioning and study how a sensory stimuli or event changes the brain's activity. We will use ERPs to look at the neural correlates underlying the process of literal and nonliteral interpretations of an utterance. Specifically, in the present study we examined three main known language ERP components, including the P2, a positive waveform observed between 150 and 250ms after word onset which typically has a frontal distribution and indexes visual processing of word form, the N400, a negative waveform observed between 300 and 500 ms after word onset which typically has a centro-parietal distribution and indexes word-level meaning, and P600, a positive waveform observed between 600 and 900ms after word onset which typically has a centro-parietal distribution and indexes the reinterpretation due to syntactic/semantic anomaly (Coulson & Lovette, 2010).

The present study aimed to record the ERP components from healthy adult participants during the comprehension of a statement after listening to a scenario vignette. Participants were first given an audio recording of a scenario vignette describing the context of a story, then they were asked to read a statement that follows the story and was intended either sincerely or sarcastically. ERPs will be recorded with the onset of the final word of the statement, which is designed to be the critical word that determines the sincerity of the sentence while also elicit either positive or negative emotions.

According to the Gricean model, we would predict to see a main effect of sincerity such that sarcastic utterances would elicit larger amplitude in N400, which indexes the unexpected nature of the sarcastic word, and a larger amplitude P600, which indexes the inferential operations that would be more pronounced for the sarcastic utterances than the sincere ones. In this model, we might expect to also see an effect of word valence but is independent of the sincerity effect, so there would be no interaction between the two variables. The direct access model would predict that the N400 effect would be the same for sincere and sarcastic statements, and there would be an opposite valence effect in sarcastic and sincere statements. There would be an interaction between sincerity and valence because sincere negative words would elicit a similar brain response as sarcastic positive words and sincere positive words would elicit a similar brain response as sarcastic negative ones.

3. Experiment

Participants. 32 right-handed, monolingual, native English speakers at UCSD students participated in the study, including 16 male and 16 female. Participants had normal or corrected to normal vision, and no history of reading difficulties or neurological disorders.

Materials. The stimuli materials included 120 short 3-4 sentence-scenarios each ended with an utterance intended sincerely or sarcastically. The individual scenarios could be either inviting critical speech act or inviting favorable speech act. Each participant saw one of the sincere or sarcastic endings for each scenario. Multiple lists were used so that both the sincere and sarcastic completion for each scenario could be tested, while each participant read 60 sarcastic and 60 sincere endings, across participants each scenario occurred with both its sarcastic and sincere

ending (see Table 1). Materials were normed in two separate studies using participants drawn from the same subject pool as the EEG study. First, to collect cloze probabilities, participants were given the first part of the critical sentence (e.g. “Your child always makes me feel so”) and asked to complete it with the first word that came to mind. The cloze probability was the proportion of participants who completed each sentence with the critical word used in the experiment (see Table 2). Second, to understand how participants perceive utterances as sarcastic or not, participants were given the story context to read and asked to rate how sarcastic each example was on a five-point scale in which 1 was “completely literal”, 3 was “neutral”, and 5 was “completely sarcastic” (see Table 3).

Table 1. Sample stimulus pair

Scenario inviting critical speech act
The baseball game was very close, and the fans, players and coaches were all nervous about the outcome. The score was tied in the final inning and Dave's team was up to bat. The first batter had struck out, and Dave was next to bat at the plate. He hadn't hit the ball once during the entire game and he'd made some major fielding errors too. His teammate, Joe, was disgusted with his performance, especially when Dave then struck out. As Dave walked back to the bench, Joe shouted • LITERAL ENDING: With you on the team, we're guaranteed to suck. • SARCASTIC ENDING: With you on the team, we’re guaranteed to dominate.
Scenario inviting favorable remarks
Jim and Bob were avid fans of their hometown football team, and never missed a game. Sunday, their team would play its toughest rival of the season, who had remained undefeated so far. But while talking about the upcoming game, they realized that two of the rival team's star players would not be playing, as they had been hurt in the previous game and were now out for the rest of the season. Jim said he was still not sure if their team would play well. Bob laughed and said • LITERAL ENDING: Yeah, with all those injuries, we’re guaranteed to dominate. • SARCASTIC ENDING: Yeah, with all those injuries, we’re guaranteed to suck.

Table 2. Cloze probability for 4 conditions

Condition Name	**Minimum**	**Maximum**	**Standard deviation**	**Mean**
Sarcastic Critical	0	0.395	0.067	0.041

Sincere Critical	0	0.279	0.057	0.034
Sarcastic Compliment	0	0.318	0.066	0.037
Sincere Compliment	0	0.256	0.066	0.049

Table 3. Sarcastic Rating for 4 conditions

Condition Name	**Minimum**	**Maximum**	**Standard deviation**	**Mean**
Sarcastic Critical	3.15	5.00	0.451	4.505
Sincere Critical	1.00	2.77	0.498	1.765
Sarcastic Compliment	3.46	4.96	0.422	4.464
Sincere Compliment	1.00	2.81	0.471	1.791

Procedure. Participants were first measured with the WM span using the reading span task (Daneman & Carpenter, 1980). Then they are assessed how frequently they utilize sarcasm in everyday communication using the Sarcasm Self-Report survey (Ivanko, Pexman, & Olineck, 2004). The survey is previously shown to predict production of sarcasm in the lab as well as reaction time differences in reading sarcastic versus literal statements.

Participants' task was to listen to the scenario, then read the ending statement which is displayed one word at a time on the screen. After reading the statements, participants were asked to answer True/False comprehension probes where participants were asked to judge the plausibility of the statement to be sarcastic.

EEG Recording. An Electro-Cap with twenty-nine tin electrodes was used to record the electroencephalogram (EEG). It was referenced online to left mastoid, and re-referenced to the average of the right and left mastoids. An electrode was placed under the right eye to monitor blinks, and electrodes placed at the outer canthus of each eye referenced to each other were used to monitor horizontal eye movements. The recording utilized amplifiers set at a band pass of 0.01-40 Hz and was digitized at a sampling rate of 250 Hz (Coulson & Lovette, 2010).

Data Analysis. ERPs time locked to the critical (final) word in each utterance. Mean amplitude measurements were made in three time intervals after word onset: 150-250ms (P2 component), 300-500ms (N400 component), and 600-900ms (P600 component).

4. Results

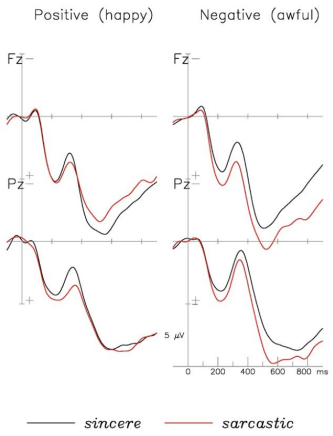

Figure 1. ERP plot in each of the four conditions as measured over two electrode sites, Fz and Pz

The data revealed that the sarcasm manipulation affected three ERP components: the P2, the N400, and the P600. It shows a difference in sarcasm effect on positive words and negative words, specifically this difference was more prominent on the negative words (figure 1). To evaluate the significance of the observed effects, we ran the repeated measures ANOVA test with factors: Sincerity, Word Valence, Laterality, and Anteriority. We analyzed the mean amplitude for each of the three ERP components separately.

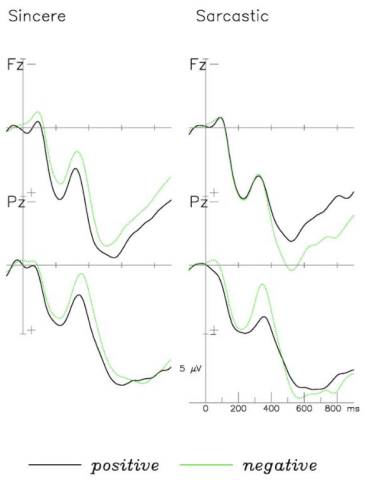

Figure 1.2 ERP amplitude showing the valence effect for sincere and sarcastic statements measured at Fz and Pz electrode sites

4.1 P2

The mean amplitude of ERPs between 150 and 250ms after word onset were analyzed with repeated measures ANOVA using factors Sincerity (sincere, sarcastic), Word Valence (positive, negative), Laterality (Left Hemisphere, Midline, Right Hemisphere), and Anteriority (Frontal, Frontocentral, Central, Centroparietal, Parietal, Occipital). The analysis revealed mean effects of Sincerity ($F(1,31)=5.06$, $p<0.05$) and Valence ($F(1,31)=7.05$, $p<0.05$). ERPs were more positive for words in sarcastic utterances than sincere ones, and positively valenced words such as "happy" elicited more positive P200 than negatively valenced words such as "awful". Although there were not significant interactions between Sincerity and Word Valence, the valence effect was larger over the frontal sites in the sincere statements (fig 2.2).

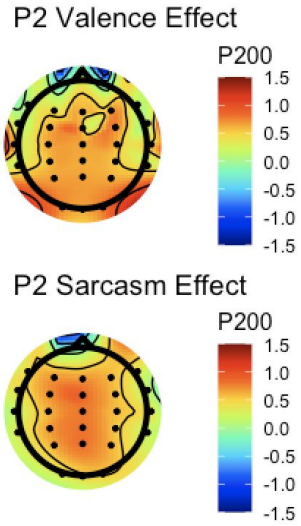

Figure 2.1 Scalp distribution of valence effect and sarcasm effect

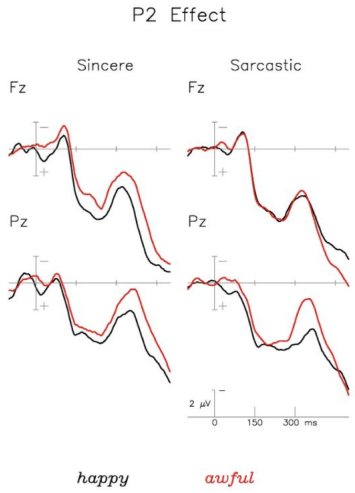

Figure 2.2 P2 effect over the frontal and parietal sites for four conditions - Sincere, Sarcastic, Negative valence, Positive valence

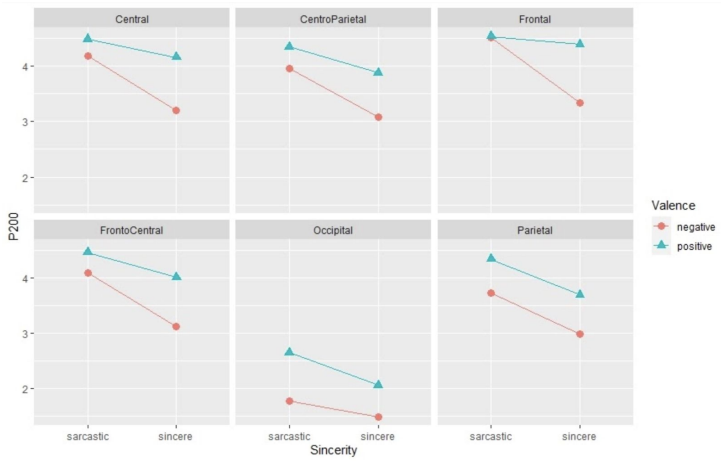

Figure 2.3 Line graph showing P2 mean amplitude with negative and positive valence at 6 electrode sites

4.2 N400

The mean amplitude of ERPs between 300 and 500ms after word onset were analyzed with repeated measures ANOVA using factors Sincerity (sincere, sarcastic), Word Valence (positive, negative), Laterality (Left Hemisphere, Midline, Right Hemisphere), and Anteriority (Frontal, Frontocentral, Central, Centroparietal, Parietal, Occipital). This analysis revealed main effects of Sincerity ($F(1,31)=5.5$, $p<0.05$), and Valence ($F(1,31) = 8.5$, $p<0.05$), as well as

interactions between Valence and Anteriority ($F(5, 155)=3.0$, $p<0.05$), and Sincerity x Valence x Anteriority ($F(5, 155)=7.4$, $p<0.05$).

The Sincerity x Valence x Anteriority interaction reflects differences in the way the valence manipulation impacted ERPs in the sincere versus the sarcastic utterances. At posterior sites where N400 is typically focused, the negatively valenced words elicited more negative N400, and this effect is similar for sarcastic and sincere conditions. In the sincere utterances, negatively valenced words like 'awful' elicit a frontally focused negativity, while in the sarcastic utterances, negatively valenced words elicit a frontally focused positivity. The valence effects are largely absent over anterior sites and the trend is in the opposite direction. This might be a reflection of a phenomenon called the negativity bias - a phenomenon in which people devote more processing resources to negatively valenced words (Vaish et al. 2013). In sincere negative conditions, a more negative response is elicited at frontal regions.

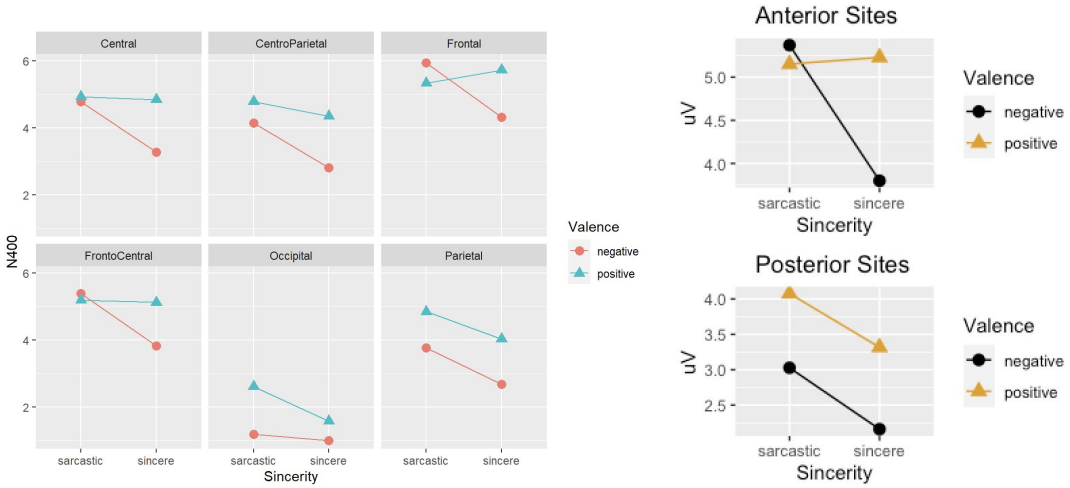

Figure 3.1 (left) line graph showing the mean N400 amplitude for 6 electrode sites (right) the average of N400 amplitude at frontal, fronto-central, and central regions are taken to plot the line graph for Anterior sites, and the average of centro-parietal, parietal, and occipital are taken to plot the line graph for Posterior sites.

4.3 P600

The mean amplitude of ERPs between 600 and 900ms after word onset were analyzed with repeated measures ANOVA using factors Sincerity (sincere, sarcastic), Word Valence (positive, negative), Laterality (Left Hemisphere, Midline, Right Hemisphere), and Anteriority

(Frontal, Frontocentral, Central, Centroparietal, Parietal, Occipital). This analysis revealed no main effect of Sincerity or Valence, but interaction effects between Sincerity and Valence ($F(1, 31)=4.9$, $p<0.05$), and Sincerity x Valence x Anteriority ($F(25, 775)=3.3$, $p<0.05$). Interactions were seen at the central, centro-parietal, frontal, fronto-central, and parietal regions, with the largest interaction at the frontal region. The interaction reflects that negatively valenced words such as 'awful' elicited more positive ERPs when they occurred in a sarcastic utterance, especially over frontal electrode sites. Positively valenced words such as 'happy' elicited more negative ERPs when they occurred in a sarcastic utterance (figure 4.1).

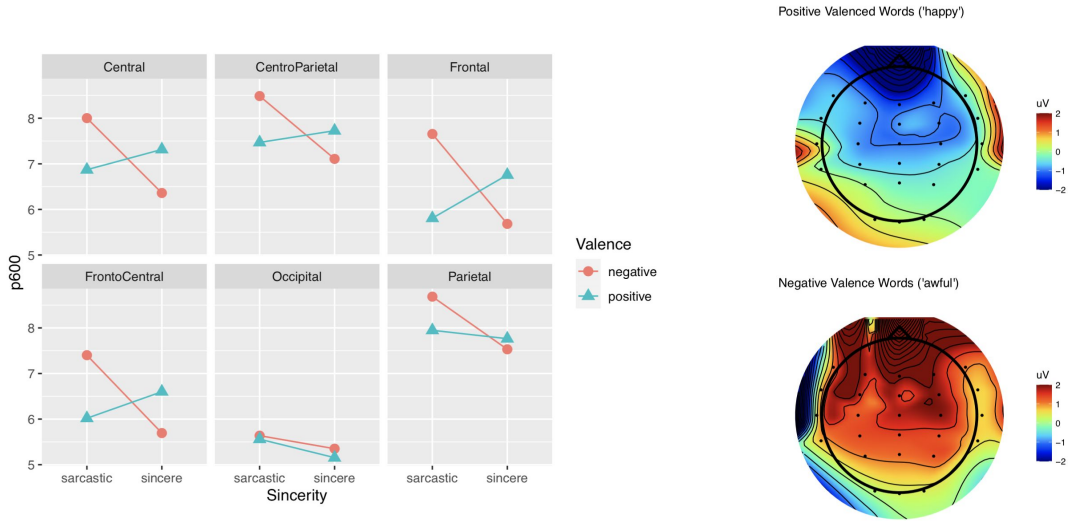

Figure 4.1 (left) Line graph showing P2 mean amplitude with negative and positive valence at 6 electrode sites. (right) scalp distributions for positive and negative valenced words in sarcastic utterances.

5. Discussion

In sum, the sarcasm manipulation affected three ERP components: the P2, the N400, and the P600. The results argue against both the Gricean model and Direct access model, as our results did not match the hypothesized ERP effects according to the two models. The present study suggests that neither of the two models are accurate, and that literal and nonliteral meaning could be processed at the same time over different portions of the brain. The pattern of data observed for the P2 matches the prediction by the Gricean model. The larger P2 for sarcastic utterances suggests that the brain registers contextual incongruity of the word with the larger context. The larger P2 amplitude for positively valenced words suggests that the brain

discriminates words associated with good things from bad things. The lack of interaction between Sincerity and Valence suggests that during this time window 150 - 150 ms post-onset, these two processes are independent of one another. In addition, the P2 component is typically believed to index visual processing of word form. The P2 component elicited by written words is typically assumed to reflect high level aspects of the visual processing of the stimuli rather than accessing the meaning of the words. Thus, its amplitude is a function of participants' expectations regarding particular words.

To evaluate the predictions made according to the Gricean model, we suggest that there is evidence against the model. There was no main effect of sarcasm in the N400 interval. Previously, we hypothesized that there would be a large N400 for sarcasm which reflects semantic anomaly, however, the effect is not seen in our analysis. Furthermore, there was also evidence against the direct access model. First, the P2 effects suggest that sarcasm and valence have independent effects on the ERPs. This is predicted by the Gricean model, but not the direct access view which suggests sarcasm will interact with valence effects. Second, the posterior electrode sites during the N400 interval show similar valence effects during the sincere and the sarcastic utterances suggests that part of the brain is accessing the literal meanings that give rise to those valence effects. This implies that different brain regions (the anterior and posterior regions) might be processing literal and nonliteral meanings simultaneously instead of having a set order.

In conclusion, the findings of the present study do not match either of the two existing models. Our data suggests that people do not either process the literal meaning first and access the nonliteral meaning according to contextual appropriateness, or directly access literal meaning, but people may process literal and nonliteral meaning at the same time separately at the anterior and posterior brain regions. Possibly there are individual differences in how people process sarcastic utterances. Future research could investigate more into factors that give rise to individual differences of how people process sarcastic utterances.

Acknowledgments

I would like to thank Dr. Seana Coulson, Dr. Megan Bardolph, Crystal Poole, Geoffrey Lizar, and members at the Brain Cognition Lab for supporting me through the entire process working on this project.

References

Coulson Seana & Lovett Christopher 2010. Comprehension of non-conventional indirect requests: An event-related brain potential study. *Italian Journal of Linguistics* 22.1.

Clark Herbert H. 1979. Responding to indirect speech acts. *Cognitive Psychology* 11. 430-477.

Clark Herbet H. & Peter Lucy 1975. Understanding what is meat from what is said: A study in conversationally conveyed requests. *Journal Verbal Learning and Verbal Behavior* 14. 56-72.

Gibbs Raymond W. 1994. T*he poetics of mind: Figurative thought, language, and understanding.* New York: Cambridge University Press.

Vaish Amrisha, Grossmann Tobias & Woodward Amanda 2013. Not all emotions are created equal: The negativity bias in social-emotional development. *Psychol Bull*. 2008 May; 134(3): 383–403.